

Attorney Docket No.: 015358-007300US

Client Reference No.: ID-RSV-263A

NON-PROVISIONAL

PATENT APPLICATION

TECHNIQUES FOR ANNOTATING MULTIMEDIA INFORMATION

Inventor(s): Jonathan J. Hull, a citizen of The United States, residing at
751 Laurel Street PMB 434
San Carlos, CA 94070

Jamey Graham, a citizen of The United States, residing at
1196 Shasta Avenue
San Jose, CA 95126

Assignee: Ricoh Company, Ltd.
15-5 Minami Aoyama 1-Chome
Minato-Ku
Tokyo, JAPAN 107-8544

Entity: Large

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 650-326-2400

TECHNIQUES FOR ANNOTATING MULTIMEDIA INFORMATION

COPYRIGHT

[01] A portion of the disclosure of this patent document contains material that is subject to copyright protection. The copyright owner has no objection to the 5 xerographic reproduction by anyone of the patent document or the patent disclosure in exactly the form it appears in the U.S. Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

CROSS-REFERENCES TO RELATED APPLICATIONS

10 [02] This is a continuation-in-part application of and claims priority from U.S. Non-Provisional Patent Application No. 09/149,921 filed September 9, 1998, the entire contents of which are herein incorporated by reference for all purposes.

5 [03] The present application also incorporates by reference for all purposes the entire contents of the following applications:

15 [04] (1) U.S. Non-Provisional Patent Application No. 08/995,616 filed December 22, 1997;

20 [05] (2) U.S. Non-Provisional Patent Application No. ____/____,____,
(Attorney Docket No. 15358-006500US) entitled "PAPER-BASED INTERFACE FOR
MULTIMEDIA INFORMATION" filed concurrently with this application;

25 [06] (3) U.S. Non-Provisional Patent Application No. ____/____,____,
(Attorney Docket No. 15358-007200US) entitled "TECHNIQUES FOR RETRIEVING
MULTIMEDIA INFORMATION USING A PAPER-BASED INTERFACE" filed
concurrently with this application;

30 [07] (4) U.S. Non-Provisional Patent Application No. ____/____,____,
(Attorney Docket No. 15358-007400US) entitled "PAPER-BASED INTERFACE FOR
MULTIMEDIA INFORMATION STORED BY MULTIPLE MULTIMEDIA
DOCUMENTS" filed concurrently with this application; and

[08] (5) U.S. Non-Provisional Patent Application No. ____/____,____,
(Attorney Docket No. 15358-007500US) entitled "TECHNIQUES FOR GENERATING A
COVERSHEET FOR A PAPER-BASED INTERFACE FOR MULTIMEDIA
INFORMATION" filed concurrently with this application.

BACKGROUND OF THE INVENTION

[09] The present invention relates to techniques for accessing multimedia information, and more particularly to techniques for generating a printable representation of the multimedia information that can be printed on a paper medium to provide a paper-based 5 interface for the multimedia information.

[10] With the rapid growth of computers, an increasing amount of information is being stored in the form of electronic (or digital) documents. These electronic documents include multimedia documents that store multimedia information. The term "multimedia information" is used to refer to information that comprises information of 10 several different types in an integrated form. The different types of information included in multimedia information may include a combination of text information, graphics information, animation information, sound (audio) information, video information, and the like. Multimedia information is also used to refer to information comprising one or more objects wherein the objects include information of different types. For example, multimedia objects included in multimedia information may comprise text information, graphics information, animation information, sound (audio) information, video information, and the like. 15

[11] Several different techniques and tools are available today for accessing and navigating multimedia information that may be stored in electronic multimedia documents. Examples of such tools and/or techniques include proprietary or customized 20 multimedia players (e.g., RealPlayer™ provided by RealNetworks, Microsoft Windows Media Player provided by Microsoft Corporation, QuickTime™ Player provided by Apple Corporation, Shockwave multimedia player, and others), video players, televisions, personal digital assistants (PDAs), and the like.

[12] The tools and techniques described above that are conventionally 25 available for accessing multimedia information focus on the electronic or digital nature/format of the multimedia information and output the multimedia information in electronic or digital form. For example, multimedia players typically execute on a computer system and output the multimedia information stored in multimedia documents via output devices coupled to the computer such as a monitor, a speaker, and the like.

[13] While retrieving multimedia information in digital form is adequate for 30 some users, it is a well-known fact that many users find it easier to comprehend and assimilate information when the information is printed on a paper medium rather than in the digital form. These users thus prefer to access information in a paper format by printing the

information on a paper medium. For example, most people who encounter a long document will typically print the document on paper before reading the document, even though there are several tools (e.g., word processors, browsers, etc.) available for viewing and navigating the document in electronic form. While there are several tools available for printing ordinary
5 data files containing text and images on paper (e.g., a printer coupled to a word-processor), there are no techniques or tools that allow users to print multimedia information on a paper-medium in a format and style that is readable by the user. As described above, all of the conventionally available tools and techniques for retrieving multimedia information focus on the electronic or digital nature/format of the multimedia content and output the multimedia
10 information in electronic or digital form.

[14] In light of the above, there is a need for techniques that allow users to access multimedia information via a paper-based interface.

BRIEF SUMMARY OF THE INVENTION

[15] The present invention provides techniques for generating a printable representation of multimedia information that can be printed on a paper medium to provide a paper-based interface for the multimedia information. According to the teachings of the present invention, the printable representation for the multimedia information may be annotated to identify locations of information in the multimedia information that may be of interest to a user. A multimedia paper document generated by printing the annotated
20 printable representation on a paper medium displays the annotations. The annotations provide visual indications of information relevant to the user. The multimedia paper document generated according to the teachings of the present invention provides a convenient tool that allows a user to readily locate portions of the multimedia paper document that are
25 relevant to the user. Since the multimedia paper document comprises a printable representation of multimedia information, the multimedia paper document generated according to the teachings of the present invention allows the user to identify portions of multimedia information that are of interest to the user.

[16] According to an embodiment of the present invention, techniques are provided for generating a paper document for an electronically stored multimedia document storing multimedia information. The multimedia information stored by the multimedia document may include video information. In this embodiment, the present invention receives user input identifying a first concept of interest. The multimedia information stored by the multimedia document is analyzed to identify information relevant to the first concept of
30

interest. The multimedia information is printed on a paper medium to generate the paper document comprising one or more printed pages such that the information that is identified to be relevant to the first concept of interest is annotated when printed on the one or more pages.

[17] According to another embodiment of the present invention, a paper document is provided that comprises one or more pages, wherein at least one page of the one or more pages is imprinted with text information extracted from multimedia information stored electronically in a multimedia document and imprinted with one or more video frames extracted from the multimedia information stored by the multimedia document. Portions of the text information printed on the at least one page and related to a topic of interest are annotated.

[18] The foregoing, together with other features, embodiments, and advantages of the present invention, will become more apparent when referring to the following specification, claims, and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[19] Fig. 1 is a simplified block diagram of a distributed system that incorporates an embodiment of the present invention;

[20] Fig. 2A depicts a networked system including a multifunction device according to an embodiment of the present invention;

[21] Fig. 2B depicts a user interface that is displayed to the user by a multifunction device according to an embodiment of the present invention;

[22] Fig. 3 is a simplified block diagram of a computer system according to an embodiment of the present invention.;

[23] Fig. 4 is a simplified high-level flowchart depicting a method of generating a printable representation of multimedia information according to an embodiment of the present invention;

[24] Figs. 5A and 5B depict a sample template according to an embodiment of the present invention;

[25] Fig. 6 is a simplified high-level flowchart depicting processing performed in step 408 of Fig. 4 according to an embodiment of the present invention;

[26] Fig. 7A depicts a page from a multimedia paper generated according to an embodiment of the present invention for a multimedia document;

[27] Fig. 7B depicts a second page that follows the page depicted in Fig. 7A in a multimedia paper document according to an embodiment of the present invention;

[28] Fig. 7C depicts a page from a multimedia paper generated according to an embodiment of the present invention for a multimedia document;

[29] Fig. 8A, 8B, and 8C depict pages from a multimedia paper document generated for a recorded meeting according to an embodiment of the present invention;

5 [30] Figs. 9A, 9B, and 9C depict pages of a multimedia paper document displaying visual markers to denote various attributes of the audio information or of the CC text information included in the multimedia information for the multimedia document for which the multimedia paper document is generated according to an embodiment of the present invention;

10 [31] Fig. 10 depicts a page from a multimedia paper document whose contents have been annotated according to an embodiment of the present invention;

[32] Fig. 11 depicts a user profile that may be configured by a user according to an embodiment of the present invention to facilitate selection of keyframes relevant to user-specified topics of interest;

15 [33] Fig. 12 depicts modules that facilitate selection of keyframes relevant to topics of interest according to an embodiment of the present invention;

[34] Fig. 13A is a simplified high-level flowchart depicting a method of accessing multimedia information using a multimedia paper document according to an embodiment of the present invention;

20 [35] Fig. 13B is a simplified high-level flowchart depicting a method of accessing multimedia information from a particular time point using a multimedia paper document according to an embodiment of the present invention;

25 [36] Fig. 14 is a simplified high-level flowchart depicting a method of generating a single printable representation according to an embodiment of the present invention that includes multimedia information selected from a plurality of multimedia documents by analyzing the printable representations of the plurality of multimedia documents;

30 [37] Fig. 15 is a simplified high-level flowchart depicting another method of generating a single printable representation that includes information extracted from a plurality of multimedia documents by analyzing the multimedia information stored by the plurality of multimedia documents according to an embodiment of the present invention;

[38] Figs. 16A, 16B, 16C, and 16D depict pages of a multimedia paper document generated according an embodiment of the present invention using the method depicted in Fig. 14;

[39] Fig. 17 depicts a coversheet generated for a multimedia paper document according to an embodiment of the present invention;

[40] Fig. 18 depicts a coversheet generated for a multimedia paper document according to another embodiment of the present invention;

5 [41] Fig. 19 depicts a coversheet generated according to another embodiment of the present invention for a multimedia paper document that has been annotated based upon user-specified topics of interest;

[42] Fig. 20 depicts a coversheet generated according to an embodiment of the present invention for a multimedia paper document that includes pages selected from
10 multiple multimedia paper documents based upon selection criteria;

[43] Fig. 21 depicts another coversheet generated according to an embodiment of the present invention for a multimedia paper document that includes pages selected from multiple multimedia paper documents based upon selection criteria; and

[44] Fig. 22 depicts a coversheet generated according to an embodiment of the present invention for a multimedia paper document that has been generated for a recorded meeting.

DETAILED DESCRIPTION OF THE INVENTION

[45] The present invention provides techniques for generating a printable representation of multimedia information that can be printed on a paper medium to provide a paper-based interface for the multimedia information. The paper-based interface provided by the present invention provides a more readable and comprehensible representation of the multimedia information.

[46] According to an embodiment of the present invention, the printable representation for the multimedia information may be annotated to identify locations of information in the multimedia information that may be of interest to a user. A paper document generated by printing the annotated printable representation on a paper medium displays the annotations. The annotations provide visual indications of information relevant to the user. For example, information printed in the paper document that is relevant to topics of interest specified by a user may be annotated or highlighted. In this manner, the multimedia paper document generated according to the teachings of the present invention provides a convenient tool that allows a user to readily locate portions of the paper document that are relevant to the user. Since the multimedia paper document comprises a printable representation of multimedia information, the paper document generated according to the

teachings of the present invention allows the user to identify portions of multimedia information that are of interest to the user.

[47] According to an embodiment of the present invention, the paper document generated by printing the printable representation on a paper medium also provides an interface for accessing or retrieving multimedia information in electronic form. The paper document may thus be used as an indexing and retrieval tool for retrieving multimedia information. For example, a user may use a paper document generated for a video recording to access or retrieve portions of the video recording.

[48] According to an embodiment of the present invention, the present invention provides techniques for generating a single printable representation that includes multimedia information extracted from a plurality of different multimedia documents or multimedia sources. According to an embodiment of the present invention, the single printable representation includes multimedia information selected from the plurality of multimedia documents based upon selection criteria. A user may specify the selection criteria. The selection criteria may be based upon any attributes of the multimedia documents or their contents, or upon user-specified topics of interest, and the like. The single or consolidated printable representation can then be printed on a paper medium to generate a consolidated paper document comprising information that satisfies the selection criteria.

[49] According to an embodiment of the present invention, the present invention provides techniques for generating a coversheet for a paper document generated by printing the printable representation on a paper medium. The coversheet may provide a summary of the contents printed on pages of the paper document.

[50] As described above, the printable representation of the multimedia information can be printed on a paper medium to generate the paper-based interface. The term "paper" or "paper medium" as used in this application is intended to refer to any tangible medium on which information can be printed, written, drawn, imprinted, embossed, etc. For purposes of this invention, the term "printing" is intended to include printing, writing, drawing, imprinting, embossing, and the like. For purposes of this invention, the document generated by printing the printable representation on a paper medium will be referred to as "multimedia paper" or "multimedia paper document." The multimedia paper document takes advantage of the high resolution and portability of paper and provides a readable representation of the multimedia information. According to the teachings of the present invention, a multimedia paper document may also be used to select, retrieve, and access the multimedia information.

[51] The multimedia information for which the multimedia paper document is generated may be stored in an electronic multimedia document. Accordingly, the term “multimedia document” is intended to refer to any storage unit (e.g., a file) that stores multimedia information in digital format. Various different formats may be used to store the multimedia information. These formats include various MPEG formats (e.g., MPEG 1, MPEG 2, MPEG 4, MPEG 7, etc.), MP3 format, SMIL format, HTML+TIME format, WMF (Windows Media Format), RM (Real Media) format, Quicktime format, Shockwave format, various streaming media formats, formats being developed by the engineering community, proprietary and customary formats, and others. Examples of multimedia documents include video recordings, MPEG files, news broadcast recordings, presentation recordings, recorded meetings, classroom lecture recordings, broadcast television programs, and the like.

[52] As previously described, multimedia information comprises information of different types in an integrated form. For example, multimedia information may comprise a combination of text, graphics, animation, sound (audio), and/or video information in an integrated form. For example, a video recording of a television broadcast may comprise video information and audio information. In certain instances the video recording may also comprise close-captioned (CC) text information which comprises material related to the video information, and in many cases, is an exact representation of the speech contained in the audio portions of the video recording. As another example, a recording of a presentation may store information captured during a presentation including video information, audio information, CC text information, information corresponding to slides presented during the presentation, whiteboard information, and other types of information. As described below, the present invention generates a printable representation of the multimedia information that includes printable representations of the various types of information included in the multimedia information. The printable representation of the multimedia document can then be printed on a paper medium to generate a multimedia paper or multimedia paper document for the multimedia information stored by the multimedia document.

30 [53] GENERATING PRINTABLE REPRESENTATION OF MULTIMEDIA INFORMATION

[54] As described above, according to an embodiment of the present invention, techniques are provided for generating a printable representation of multimedia information that can be printed on a paper medium to produce a multimedia paper document.

The multimedia paper document provides a paper-based interface for the user to view and comprehend the multimedia information. Fig. 1 is a simplified block diagram of a distributed system 100 that might incorporate an embodiment of the present invention. As depicted in Fig. 1, distributed system 100 comprises a number of devices or computer systems including one or more user systems 102, a multimedia information processing server system (MIPSS) 104, a multimedia information source (MIS) 106, and a multimedia paper output device 108 coupled to communication network 110 via a plurality of communication links. It should be apparent that distributed system 100 depicted in Fig. 1 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. For example, in alternative embodiments of the present invention, one or more of the systems depicted in Fig. 1 (e.g., MIPSS 104 and output device 108) may be incorporated into a single system. In other alternative embodiments, the present invention may also be embodied in a stand-alone system, and the like.

[55] Communication network 110 provides a mechanism allowing the various devices and computer systems depicted in Fig. 1 to communicate and exchange data and information with each other. Communication network 110 may itself be comprised of many interconnected computer systems and communication links. While in one embodiment, communication network 110 is the Internet, in other embodiments, communication network 110 may be any suitable communication network including a local area network (LAN), a wide area network (WAN), a wireless network, an intranet, a private network, a public network, a switched network, and the like.

[56] The communication links used to connect the various systems depicted in Fig. 1 may be of various types including hardwire links, optical links, satellite or other wireless communications links, wave propagation links, or any other mechanisms for communication of information. Various communication protocols may be used to facilitate communication of information via the communication links. These communication protocols may include TCP/IP, HTTP protocols, extensible markup language (XML), wireless application protocol (WAP), protocols under development by industry standard organizations, vendor-specific protocols, customized protocols, and others.

[57] According to the teachings of the present invention, MIPSS 104 is configured to perform processing to facilitate generation of a printable representation of the multimedia information. The printable representation generated by MIPSS 104 for a multimedia document may include printable representations of the various types of

information included in the multimedia information stored by the multimedia document. The printable representation generated by MIPSS 104 may be printed on a paper medium to generate a multimedia paper document. The processing performed by MIPSS 104 to generate a printable representation of the multimedia information may be implemented by software modules executing on MIPSS 104, by hardware modules coupled to MIPSS 104, or combinations thereof. According to alternative embodiments of the present invention, the processing may also be distributed between other computer systems and devices depicted in Fig. 1.

[58] The multimedia information for which MIPSS 104 generates a printable representation may be stored in a multimedia document accessible to MIPSS 104. For example, the multimedia document may be stored by MIPSS 104 or may alternatively be stored in locations accessible to MIPSS 104.

[59] In alternative embodiments of the present invention, instead of being stored in a multimedia document, MIPSS 104 may receive a stream of multimedia information (e.g., a streaming media signal, a cable signal, etc.) from a multimedia information source such as MIS 106. Examples of MIS 106 include a television broadcast receiver, a cable receiver, a TIVO box, and the like. MIPSS 104 may receive the multimedia information directly from MIS 106 or may alternatively receive the information via a communication network such as communication network 106. MIPSS 104 may then store the multimedia information received from MIS 106 in the form of a multimedia document and use the stored information to generate the printable representation of the multimedia information.

[60] After generating the printable representation of the multimedia information, MIPSS 104 may communicate the printable representation to output device 108 that is capable of generating a multimedia paper document by printing the printable representation on a paper medium. In one embodiment, MIPSS 104 may itself be configured to generate a multimedia paper document from the printable representation of the multimedia information. In alternative embodiments, the printable representation generated by MIPSS 104 may be stored for later use.

[61] As described above, multimedia information source (MIS) 106 represents a source of multimedia information. According to an embodiment of the present invention, MIS 106 may store multimedia documents that are accessed by MIPSS 104. In alternative embodiments, MIS 106 may provide a multimedia information stream to MIPSS 104. For example, MIS 106 may be a television receiver/antenna providing live television

feed information to MIPSS 104. MIS 106 may be a video recorder providing the recorded video and/or audio stream to MIPSS 104. In alternative embodiments, MIS 106 may be a presentation or meeting recorder device that is capable of providing a stream of the captured presentation or meeting information to MIPSS 104. MIS 106 may also be a receiver (e.g., a satellite dish or a cable receiver) that is configured to capture or receive (e.g., via a wireless link) multimedia information from an external source and then provide the captured multimedia information to MIPSS 104 for further processing.

[62] Users may use user systems 102 to interact with the other systems depicted in Fig. 1. For example, a user may use user system 102 to select one or more 10 multimedia documents and request MIPSS 104 to generate multimedia papers for the selected documents. Users may also use user systems 102 to view digital versions of the multimedia documents. For example, multimedia players executing on a user system may play multimedia information stored by a multimedia document. A user system 102 may be of different types including a personal computer, a portable computer, a workstation, a computer terminal, a network computer, a mainframe, a kiosk, a personal digital assistant (PDA), a communication device such as a cell phone, or any other data processing system.

[63] Output device 108 is capable of generating a multimedia paper 20 document based upon the printable representation of the multimedia information received from MIPSS 104. Accordingly, output device 108 represents any device that is capable of outputting (e.g., printing, writing, drawing, imprinting, embossing, etc.) the printable representation of the multimedia information on a paper medium. For example, output device 108 may be a printer that is coupled to MIPSS 104. The printer may be configured to receive a signal from MIPSS 104 including a printable representation of multimedia 25 information from MIPSS 104, and to generate a multimedia paper document based upon the printable representation of the multimedia information.

[64] According to an embodiment of the present invention, output device 108 may be incorporated as part of a multi-function device (or MFD) that is capable of performing a plurality of different functions in addition to allowing users to generate multimedia paper documents. For example, a MFD may allow users to copy, fax, or scan 30 documents including multimedia paper documents. A MFD may also allow users to perform other functions. A MFD may also allow users to select multimedia documents for which printable representations are to be generated according to the teachings of the present invention. For example, a MFD may provide a user interface that allows a user to select one or more multimedia documents, request generation of printable representations for the

selected multimedia documents, generate multimedia paper documents for the selected multimedia documents, and perform other functions such as copying, faxing, etc. on the printable representations or on the multimedia papers.

[65] Fig. 2A depicts a networked system including a MFD 200 according to an embodiment of the present invention. In the embodiment depicted in Fig. 2A, MFD 200 is coupled to MIPSS 104 that in turn is coupled to MIS 106. In the embodiment depicted in Fig. 2A, MIS 106 is a satellite dish or TV antenna that receives and provides multimedia information to MIPSS 104. MIPSS 104 generates a printable representation for the multimedia information. The printable representation may be forwarded to MFD 200 or may alternatively be stored by MIPSS 104.

[66] In the embodiment depicted in Fig. 2A, MFD 200 provides a user interface 202 that can be used by users to provide instructions or commands to MFD 200 and to view information output by MFD 200. Interface 202 comprises an area 204 that displays a list of documents 204-a including multimedia documents that can be selected by a user. The multimedia documents displayed in area 204 may be stored by MFD 200 or may be stored by other devices (such as MIPSS 104) coupled to MFD 200. In alternative embodiments, area 204 may display a list of documents accessible to MIPSS 104. The multimedia documents displayed in area 204 may correspond to television broadcast recordings, video clips, recorded meetings, etc.

[67] Area 204 may also display various details about the multimedia documents that are listed. In the embodiment depicted in Fig. 2A, for each multimedia document listed in area 204, the information displayed in area 204 includes information related to the date 204-c and time 204-b of the multimedia document recording. If a printable representation has already been generated for a multimedia document, the number of pages 204-d needed to print the printable representation (i.e., the number of pages in the multimedia paper document for the multimedia document) are also displayed. For example, the multimedia document titled "CNN/fn" stores a recording that was recorded on May 21, 2001 between 11:01AM and 1:00PM. A printable representation has been generated for the "CNN/fn" multimedia document and comprises 26 pages.

[68] A user may select one or more documents displayed in area 204 using an input device of MFD 104. In the embodiment depicted in Fig. 2A, the user may select a document by clicking on the document name in area 204 or alternatively by using "Select" button 206. For example, as shown in Fig. 2A, the user has selected a multimedia document titled "NewsHour" which corresponds to a news broadcast recorded on May 18, 2001

between 6:00-7:00PM. The user may then request MFD 200 to perform one or more functions provided by MFD 200 on the selected document(s). According to an embodiment of the present invention, the user may request generation of printable representations for the selected multimedia documents or may request generation of multimedia paper documents
5 for the selected multimedia documents. The multimedia documents displayed in area 204 may also be indexed by MFD 200 that allows a user to perform familiar operations such as keyword searching, browsing for similar documents, etc. on the selected multimedia documents.

[69] User interface 202 provides a plurality of user interface features that
10 allow a user to specify functions or operations to be performed on the selected document(s). For example, the user may select “Print” button 208 to instruct MFD 200 to print multimedia paper documents 210 for the multimedia documents selected by the user in area 204.
15 According to an embodiment of the present invention, upon receiving a signal indicating selection of “Print” button 206 by a user, MFD 200 sends a signal to MIPSS 104 requesting generation of printable representations for the user-selected multimedia documents. If
20 printable representations for the user-selected documents already exist, MIPSS 104 communicates the previously generated printable representations for the user-selected documents to MFD 200. Alternatively, if the printable representations do not exist, MIPSS 104 generates printable representations for the one or more user-selected documents and then provides the printable representations to MFD 200. MFD 200 may then generate (or print) multimedia paper documents for the user-selected documents based upon printable representations corresponding to the documents received from MIPSS 104. In alternative embodiments, MFD 200 may itself be configured to generate printable representations for user-selected multimedia documents.

[70] User interface 202 also provides a “Play” button 212 which when selected by a user causes MFD 200 to playback multimedia information from the user-selected multimedia document(s). For example, Fig. 2B depicts a user interface 214 that is displayed to the user upon selection of “Play” button 212 according to an embodiment of the present invention. Interface 214 allows the user to play back video and audio information contained in the “NewsHour” multimedia document selected by the user in area 204 of Fig.
25 30 2A. If MFD 200 is connected to one or more output devices (e.g., an output monitor, other networked output devices), the user may also select the output device to be used for the playback. For example, the user may indicate that the information is to be played back on the user’s computer in the user’s office (or on a television in a particular conference room, etc.).

In specific embodiments of the present invention, the user may also indicate the time when the multimedia information is to be played back.

[71] Referring back to Fig. 2A, user interface 202 also provides a numeric keypad 216 that facilitates operations such as faxing of documents. For example, using 5 keypad 216, a user may fax a multimedia paper document or a printable representation of a user-selected multimedia document to a recipient. The user may also make copies of the multimedia paper document by selecting “Copy” button 218. “Cancel” button 220 allows the user to cancel a pre-selected function.

[72] It should be apparent that MFD 200 and the user interfaces depicted in 10 Figs. 2A and 2B are merely illustrative of an embodiment incorporating the present invention and do not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. For example, in a networked environment, a web browser-enabled interface may be provided allowing a user to control the functions of MFD 200 from a remote location, for example, using the user’s computer system or PDA, and the like.

[73] Fig. 3 is a simplified block diagram of a computer system 300 according to an embodiment of the present invention. Computer system 300 may be used as any of the computer systems depicted in Fig. 1. As shown in Fig. 3, computer system 300 includes at least one processor 302 that communicates with a number of peripheral devices via a bus subsystem 304. These peripheral devices may include a storage subsystem 306, comprising a memory subsystem 308 and a file storage subsystem 310, user interface input devices 312, user interface output devices 314, and a network interface subsystem 316. The input and output devices allow user interaction with computer system 300. A user may be a human user, a device, a process, another computer, and the like. Network interface 20 25 subsystem 316 provides an interface to other computer systems and communication networks including communication network 110.

[74] Bus subsystem 304 provides a mechanism for letting the various components and subsystems of computer system 300 communicate with each other as intended. The various subsystems and components of computer system 300 need not be at 30 the same physical location but may be distributed at various locations within network 100. Although bus subsystem 304 is shown schematically as a single bus, alternative embodiments of the bus subsystem may utilize multiple buses.

[75] User interface input devices 312 may include a keyboard, pointing devices, a mouse, trackball, touchpad, a graphics tablet, a scanner, a barcode scanner, a

touchscreen incorporated into the display, audio input devices such as voice recognition systems, microphones, and other types of input devices. In general, use of the term “input device” is intended to include all possible types of devices and ways to input information using computer system 300.

5 [76] User interface output devices 314 may include a display subsystem, a printer, a fax machine, or non-visual displays such as audio output devices. The display subsystem may be a cathode ray tube (CRT), a flat-panel device such as a liquid crystal display (LCD), or a projection device. In general, use of the term “output device” is intended to include all possible types of devices and ways to output information from computer system
10 300.

13 [77] Storage subsystem 306 may be configured to store the basic programming and data constructs that provide the functionality of the computer system and of the present invention. For example, according to an embodiment of the present invention, software modules implementing the functionality of the present invention may be stored in storage subsystem 306 of MIPSS 104. For example, software modules that facilitate generation of printable representations of the multimedia information may be stored in storage subsystem 306 of MIPSS 104. These software modules may be executed by processor(s) 302 of MIPSS 104. In a distributed environment, the software modules may be stored on a plurality of computer systems and executed by processors of the plurality of computer systems. Storage subsystem 306 may also provide a repository for storing various databases and files that may be used by the present invention. For example, the multimedia documents may be stored in storage subsystem 306. Storage subsystem 306 may comprise memory subsystem 308 and file storage subsystem 310.
20

25 [78] Memory subsystem 308 may include a number of memories including a main random access memory (RAM) 318 for storage of instructions and data during program execution and a read only memory (ROM) 320 in which fixed instructions are stored. File storage subsystem 310 provides persistent (non-volatile) storage for program and data files, and may include a hard disk drive, a floppy disk drive along with associated removable media, a Compact Digital Read Only Memory (CD-ROM) drive, an optical drive, removable media cartridges, and other like storage media. One or more of the drives may be located at remote locations on other connected computers.
30

[79] Computer system 300 itself can be of varying types including a personal computer, a portable computer, a workstation, a computer terminal, a network computer, a mainframe, a kiosk, a personal digital assistant (PDA), a communication device

such as a cell phone, a game controller, or any other data processing system. Due to the ever-changing nature of computers and networks, the description of computer system 300 depicted in Fig. 3 is intended only as a specific example for purposes of illustrating the preferred embodiment of the computer system. Many other configurations of a computer system are
5 possible having more or fewer components than the computer system depicted in Fig. 3. For example, several other subsystems may be included in computer system 300 depending upon the functions performed by system 300.

[80] Fig. 4 is a simplified high-level flowchart 400 depicting a method of generating a printable representation of multimedia information according to an embodiment 10 of the present invention. The processing depicted in Fig. 4 may be performed by MIPSS 104 (e.g., by software modules executing on MIPSS 104). In alternative embodiments of the present invention, the processing may be distributed among the various systems depicted in Fig. 1. Flowchart 400 depicted in Fig. 4 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[81] As depicted in Fig. 4, according to an embodiment of the present invention, the method is initiated when MIPSS 104 receives a signal requesting generation of a printable representation for a multimedia document storing multimedia information (step 20 402). Alternatively, the signal received in step 402 may request generation of a multimedia paper document for a multimedia document. MIPSS 104 may receive the signal from a variety of different sources including a user system 102, a MFD 200, from an interface provided by MIPSS 104, from MIS 106, and the like. The signal may identify the multimedia document for which a printable representation is to be generated.

[82] In alternative embodiments of the present invention, the signal received in step 402 may comprise a stream of multimedia information (e.g., from MIS 106) for which a printable representation (or multimedia paper document) is to be generated. If the signal includes a multimedia information stream, MIPSS 104 may store the stream in a multimedia document and then generate a printable representation for the document. For 30 purposes of explaining the processing in Fig. 4, it is assumed that the signal received in step 402 identifies a multimedia document for which a printable representation is to be generated.

[83] MIPSS 104 then accesses the multimedia document identified by the signal received in step 402 (step 404). The multimedia document identified by the signal received in step 402 may be stored by MIPSS 104 or may alternatively be stored by other

devices or systems from where it can be accessed by MIPSS 104. In alternative embodiments of the present invention, the signal received in step 402 may itself comprise the multimedia document.

[84] MIPSS 104 then determines layout and format information to be used for generating the printable representation (step 406). The layout and format information specifies how the information is to be printed on the paper medium. For example, the layout and format information may comprise information identifying the paper-medium and size of the paper (e.g., letter size, legal size, A4 size, etc.) for which the printable representation is to be generated. The layout and format information may also identify special features of the paper (e.g., a paper with a letterhead, paper of a particular color, etc.) for which the printable representation is to be generated. In specific embodiments of the present invention, a default paper medium (e.g., letter size paper) may be selected for generating the printable representation of the multimedia document.

[85] Additionally, the layout and format information indicates the layout and formatting features to be used for generating the printable representation of the multimedia information. For example, according to an embodiment of the present invention, the layout and format information specifies how each type of information (e.g., audio, video, images, text, etc.) included in the multimedia information is to be printed. For example, for each type of information included in the multimedia information, the layout and format information may identify the area (or location or section of the paper medium) on the paper medium in which the information is to be printed, and the format or styles (e.g., font type, font size, bolding, underlining, number of columns per page, size of the columns, page margins, etc.) to be used for printing the information. In embodiments of the present invention which support multiple languages, the layout and format information may also indicate the language (or languages) to be used for printing the information. MIPSS 104 uses the layout and format information to generate the printable representation.

[86] For example for text information (e.g., CC text, text transcript of audio information) included in the multimedia information, the layout and format information may specify the font type and font size to be used for printing the text information, the number of columns to be used for printing the information, the size and location of the columns, the color of the font to be used for printing (which may depend on the color of the paper for which the printable representation is to be generated), line spacing, length of each line, number of words per line, bolding and capitalization techniques, and the like. The layout and format information may also identify the language to be used for printing the text

information. For example, the layout and format information may indicate that the text is to be printed in two columns on each page with the English version in the first column and a Japanese translation of the English version in the second column.

[87] For audio information, the layout and format information may identify techniques to be used for converting the audio information to text information (i.e., techniques for generating a text transcript for the audio information), the format and styles for printing the audio transcript (which may be the same as for printing text information), and the like. For video information, the layout and format information may indicate how the video information is to be represented on paper. According to an embodiment of the present invention, the printable representation of the video information includes keyframes that are extracted from the video information. In this embodiment, the layout and format information may specify the sampling rate for extracting the keyframes, the number of keyframes that are to be extracted from the video information, the order and placement of the keyframes on the paper medium, and other like information.

[88] Likewise, for other types of information included in the multimedia information, the layout and format information specifies the manner in which the multimedia information is to be printed on the paper medium. Accordingly, the layout and format information specifies how printable representations are to be generated for each type of information included in the multimedia information stored by the multimedia document.

[89] According to an embodiment of the present invention, the layout and format information is stored in the form of templates. The templates may be customized for a particular type of paper. For example, a first template may be defined for letter size paper and a second template different from the first template may be defined for A4 size paper. It should be apparent that one or more templates may be defined for each type and size of paper. If multiple templates are provided, the user may be allowed to select a particular template to be used for generating the printable representation. According to an embodiment of the present invention, information identifying a user-selected template may be included in the signal received in step 402. Default templates may also be specified. The user may also be allowed to create new templates, and to edit and modify previously configured templates.

30 In this manner, the layout and format information is user-configurable.

[90] The goal of a template (or layout and format information in general) is to generate a printable representation which when printed on a paper medium generates a readable and comprehensible multimedia paper document. In order to create a readable version, the templates may adhere to many of the basic guidelines designed and used by the

newspaper industry. For instance, the use of special fonts, multiple columns and shorter lines of text along with line spacing, bolding and capitalization techniques, and other type-setting features used by the newspaper industry may be specified in the templates. The layout and format information thus contributes towards making the multimedia paper document more

5 readable and comprehensible.

[91] Figs. 5A and 5B depict a sample template according to an embodiment of the present invention. The template is defined using XML syntax but could easily be represented in other ways. The template is designed for use with letter-size (8.5 x 11 inch) sheets of white 24-lb. paper. As defined in the template, each sheet is configured to contain
10 one title zone, two text zones, and a video zone. The title zone specifies the zone or area of the paper medium where the title is to be printed and the manner in which the title is to be printed. The first text zone specifies the zone or area or section of the paper medium where the CC text included in the multimedia information is to be printed and the manner in which the CC text is to be printed. The second text zone specifies the zone or section of the paper
15 medium where the Japanese translation of the CC text is to be printed and the manner in which the Japanese text is to be printed. It should be apparent that in alternative embodiments of the present invention, CC text included in the multimedia information and which is a continuation of the information printed in the first text zone may be printed in the second text zone. The video zone specifies the zone or area of the paper medium where the
20 video information included in the multimedia document is to be printed and the manner in which the video is to be printed.

[92] The template information in Fig. 5A specifies that the title zone (identified by “ZONE_ID 0”) area is bounded by a rectangle whose left edge is located at a distance of 3 inches from the left margin of the page and whose right edge is located at a
25 distance of 3 inches from the right margin of the page (i.e., the rectangle is 2.5 inches wide). The top edge of the title zone rectangle is located 0.75 inches from the top margin of the page and the bottom edge of the rectangle is located 9.6 inches from the bottom margin of the page (i.e., the rectangle is 0.65 inches high). The text in the title zone is configured to be English and is to be extracted from the header of the video clip. The title is configured to be printed
30 in a 14 point, black Times font, and is centered within the title zone. The lines are to be single-spaced.

[93] The first text zone (identified by “ZONE_ID 1”) is also bounded by a rectangle whose left edge is located at a distance of 1.1 inches from the left margin of the page, whose right edge is located at a distance of 5.4 inches from the right margin of the

page, whose top edge is located 1.5 inches from the top margin of the page, and whose bottom edge is located 1.0 inch from the bottom margin of the page. The text in the first text zone is to be printed in the English language. The origin of the text to be printed is defined to be CC text included in the multimedia information. The text is to be printed in a black 10 point Garamond font. Lines are to be single-spaced. Subject changes in the closed caption (which are usually indicated in the CC text by three greater-than signs “>>>”) are to be shown by inserting a 1.5 line break and by bolding the first three words. Speaker changes (which are usually indicated in CC text by two greater-than signs “>>”) are to be shown with a single-line break with no emphasis. Annotations (that indicate words in the transcript that occur in a user’s profile) (described below in more detail) are to be shown with italicized text and blue underlining.

[94] The second text zone (identified by “ZONE_ID 2”) is also bounded by a rectangle whose left edge is located at a distance of 4.6 inches from the left margin of the page, whose right edge is located at a distance of 1.9 inches from the right margin of the page, whose top edge is located 1.5 inches from the top margin of the page, and whose bottom edge is located 1.0 inch from the bottom margin of the page. Unlike the first zone, the text in the second text zone is to be printed in Japanese. A translation source to facilitate the translation to Japanese is identified. The text is to be printed in a black 10 point AsianGaramond font. Lines are to be single-spaced. Subject changes in the closed caption text (which are usually indicated in CC text by three greater-than signs “>>>”) are to be shown by inserting a 1.5 line break and by bolding the first three words. Speaker changes (which are usually indicated in CC text by two greater-than signs “>>”) are to be shown with a single-line break with no emphasis. Annotations to words or phrases are to be shown with italicized text and blue underlining. Further details related to annotations are provided below.

[95] The video zone (identified by “ZONE_ID 3”) is also bounded by a rectangle whose left edge is located at a distance of 3.2 inches from the left margin of the page, whose right edge is located at a distance of 4.5 inches from the right margin of the page, whose top edge is located 1.5 inches from the top margin of the page, and whose bottom edge is located 1.0 inch from the bottom margin of the page. The source for the displayed data in the video zone is to be a set of keyframes that are generated by sampling the video channel of the multimedia information at a rate of 1 frame per second. Text in those frames is to be annotated by drawing a red box around it with line-width of 3-points. The keyframes are to be divided into sets of four. Each set is to be 0.4 inches wide and 0.3 inches high. The keyframes from each set are to be laid out within the video zone by sequentially

packing them into the available space. Each group of four keyframes is to be annotated with an interleaved 2-of-5 barcode 0.8 inches wide and 0.15 inches high that appears underneath the group.

[96] It should be apparent that the template depicted in Figs. 5A and 5B is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[97] Referring back to Fig. 4, MIPSS 104 then generates a printable representation of the multimedia information stored in the multimedia document accessed in step 402 based upon the layout and format information determined in step 406 (step 408). Generating a printable representation for the multimedia document involves generating a printable representation for each type of information included in the multimedia information based upon the layout and format information.

[98] If the signal received in step 402 requested generation of a multimedia paper document, MIPSS 104 may then print the printable representation of the multimedia information to generate the multimedia paper document (step 410). Alternatively, MIPSS 104 may communicate the printable representation of the multimedia information generated in step 408 to an output device 108 (e.g., a printer, a MFD, etc.) that is configured to generate the multimedia paper document (step 412). Other operations may also be performed on the printable representation of the multimedia information (step 414). For example, the printable representation may be stored for future generation of multimedia paper document, the information may be faxed, the information may be searched, indexed, annotated, etc., and the like.

[99] Fig. 6 is a simplified high-level flowchart depicting processing performed in step 408 of Fig. 4 according to an embodiment of the present invention. The processing depicted in Fig. 6 may be performed by software modules executing on MIPSS 104, by hardware modules coupled to MIPSS 104, or a combination thereof. In alternative embodiments of the present invention, the processing may be distributed among the various systems depicted in Fig. 1. The processing depicted in Fig. 6 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[100] As described above, in step 408 MIPSS 104 generates a printable representation of the multimedia information based upon the layout and format information

determined in step 406. As part of the processing performed in step 408, MIPSS 104 divides or indexes the multimedia information contained by the multimedia document into sequential segments or portions of a particular time length (step 602). Each segment is characterized by a starting time and an ending time. Each segment comprises multimedia information
5 occurring between the starting time and ending time associated with the segment. In other words, each segment or section comprises multimedia information for a specific time period. A sequential list of segments represents the entire multimedia information stored in the multimedia document. For example, according to an embodiment of the present invention, a 10-second time period may be used for segmenting the multimedia information. Using the
10 10-second time period value, a 5-minute video recording may be divided into 30 segments or sections. The first segment comprises multimedia information for the first 10 seconds of the multimedia document, the second segment comprises multimedia information for the next 10 seconds, the third segment comprises multimedia information for the next 10 seconds, and so on. The value of the time period to be used for segmenting the multimedia document may be
15 user-configurable.

[101] From the segments generated in step 602, MIPSS 104 then selects a set of segments or portions of the multimedia document comprising multimedia information that is to be included in the printable representation of the multimedia information (step 604). According to an embodiment of the present invention, all the segments generated in step 602 are selected to be included in the printable representation. According to other embodiments of the present invention, a subset of the segments generated in step 602 may be selected for inclusion in the printable representation of the multimedia information based upon some selection criteria. The selection criteria may be user configurable.

[102] According to one such embodiment, MIPSS 104 may compare
25 multimedia information stored by successive segments and only select those segments for inclusion in the printable representation that contain additional information relative to their preceding segment. In this manner, segments comprising repetitive or redundant information are not selected for inclusion in the printable representation. For example, there may be periods of time within a video recording wherein the audio or video content information does
30 not change (e.g., during a period of “silence” or “blankness” on the video recording). Segments comprising multimedia information corresponding to such periods of time may not be selected by MIPSS 104 in step 604.

[103] According to another embodiment of the present invention, MIPSS 104 may select only those segments for inclusion in the printable representation that contain

information relevant to the user who has requested generation of the printable representation. For example, MIPSS 104 may select only those segments for inclusion in the printable representation that contain multimedia information related to user-specified topics of interest (which may be specified in a user profile). For example, a user may have specified an
5 interest in all information related to topic “Afghanistan.” In this embodiment, MIPSS 104 may scan the multimedia information contained by the various segments and select only those segments for inclusion in the printable representation that contain information related to Afghanistan. Various techniques known to those skilled in the art may be used to facilitate selection of the segments based upon their content and their relevance to user-specified
10 topics.

15 [104] According to another embodiment of the present invention, MIPSS 104 may apply a summarization technique to select segments to be included in the printable representation. Applying the summarization technique, only those segments that satisfy some selection criteria may be selected for inclusion in the printable representation. For example, for a multimedia document corresponding to an audio recording, MIPSS 104 may only select those segments for inclusion that contain the first sentence spoken by each speaker (or alternatively segments that contain the first line of each paragraph of CC text). This reduces the size of the printable representation and as a result reduces the number of pages needed to print the printable representation. Various other techniques known to those of skill in the art
20 may also be used to determine which segments are to be included in the printable representation of the multimedia information.

25 [105] MIPSS 104 then paginates the segments (i.e., determines on which page a particular segment is to be printed) selected in step 604 (step 606). According to an embodiment of the present invention, for each page starting with the first page, MIPSS 104 determines the segments to be printed on the page based upon the layout and format
30 information which influences the amount of information that can be printed on a page. In this manner, MIPSS 104 determines the amount of multimedia information to be printed on each page and the total number of pages required to print the multimedia information stored in the multimedia document. For each page, MIPSS 104 determines the start time for information printed on the page (corresponding to the start time of the first segment printed on the page) and the end time for information printed on the page (corresponding to the end time of the last segment printed on the page).

[106] The number of segments that can be printed on a particular page is influenced by the layout and format information and the contents of the segments. The size

of the contents of each segment is in turn influenced by the time period used for segmenting the multimedia information stored in the multimedia document. For example, the amount of information stored by a segment is generally directly proportional to the value of the time period used for segmenting the multimedia document.

5 [107] According to an embodiment of the present invention, for a given template storing the layout and format information and for a particular segmentation time period, the number of segments printed on a particular page is fixed for each page of the multimedia paper document. For example, based upon a particular template and a particular time period used for segmenting the multimedia information, MIPSS 104 may determine that
10 multimedia information corresponding to "M" segments (where $M > 0$) can be printed on each page of the multimedia paper document. Based upon this segments-per-page value, MIPSS 104 may then determine the total number of pages in the multimedia paper document and the segments to be printed on each page.

15 [108] For example, for a 5-minute video recording which is divided into 30 segments using a 10-second segmentation value, and assuming that all segments are selected for inclusion in the printable representation in step 604, MIPSS 104 may determine that multimedia information corresponding to 12 segments will be printed per page of the multimedia paper document. Using this segments-per-page value, MIPSS 104 may

determine that 3 pages ($\left\lceil \frac{30}{12} \right\rceil = 3$) will be needed to print the multimedia information (i.e.,

20 the multimedia paper document will contain 3 pages). Multimedia information corresponding to segments 1-12 will be printed on the first page of the multimedia paper document, multimedia information corresponding to segments 13-24 will be printed on the second page of the multimedia paper document, and multimedia information corresponding to segments 25-30 will be printed on the last or third page of the multimedia paper document.

25 [109] In alternative embodiments of the present invention, the number of segments printed on a particular page may vary from page to page of the multimedia paper document based upon the contents of the segments. In this embodiment, the number of segments to be printed on a particular page is influenced by the type and contents of multimedia information contained in the segments selected in step 604. In this embodiment,
30 for each page, starting with the first page of the multimedia paper document, MIPSS 104 determines the number of selected segments (starting with the segment having the earliest starting time) which can be printed on each page. In this manner the number of segments that

can be printed on a page are determined on a sequential page-per-page basis starting with the first page.

[110] For example, for a 5-minute video recording which is divided into 30 segments using a 10-second segmentation value and assuming that all the segments are selected for inclusion in step 604, MIPSS 104 may determine that multimedia information corresponding to segments 1-10 can be printed on the first page, multimedia information corresponding to segments 11-25 can be printed on the second page of the multimedia paper document, and multimedia information corresponding to sections 25-30 can be printed on the third page of the multimedia paper document. Accordingly, in this embodiment of the present invention, the number of segments printed on each page of the multimedia document may vary from page to page based upon the contents of the segments to be printed. Various other techniques may also be used to determine how the selected segments will be printed.

[111] MIPSS 104 then generates a printable representation for each page determined in step 606 based upon the layout and format information (step 608). As part of step 608, for each page, MIPSS 104 determines segments associated with that page, and generates printable representations for the various types of information included in the multimedia information corresponding to the segments. Various different techniques may be used by MIPSS 104 to generate printable representation for the various types of information included in the multimedia information.

[112] For example, for CC text information included in the multimedia information, MIPSS 104 may apply the formatting styles specified by the layout and format information. For audio information, MIPSS 104 may generate a text transcript for the audio information by applying audio-to-text conversion techniques (which may also be specified in the layout and format information) and then apply the text formatting. For video information, MIPSS 104 may apply various keyframe extraction techniques (which may be specified in the layout and format information) to extract keyframes from the video information included in the selected segments of the multimedia information. According to an embodiment of the present invention, MIPSS 104 extracts keyframes that capture salient features of the video information (or keyframes that are informative) for a particular segment of the multimedia information. For example, images of faces are often quite informative. In choosing keyframes for a news broadcast, MIPSS 104 may select keyframes whose contents are different from the anchorperson. This increases the information conveyed by the keyframes.

[113] Several other techniques known to those of skill in the art may also be applied by MIPSS 104 to generate a printable representation for the multimedia information.

For example, the article “Key frame selection to represent a video,” written by Frederic Dufaux and published in the Proceedings of the International Conference on Image Processing, Vancouver, 2000, describes techniques for selecting keyframes for representing a video. The entire contents of this article are herein incorporated by reference in their entirety

5 for all purposes.

[114] The printable representation of the multimedia information generated by MIPSS in step 408 may then be printed to generate a multimedia paper document, communicated to a device capable of generating a multimedia paper document, or subjected to other operations according to step 410, 412, or 414 depicted in Fig. 4.

[115] Fig. 7A depicts a page 700 from a multimedia paper generated according to an embodiment of the present invention for a multimedia document. In the embodiment depicted in Fig. 7A, the multimedia document corresponds to a television broadcast recording. As depicted in Fig. 7A, page 700 comprises a title section 702, a first text section 704, a second text section 706, a video section 708, and a controls section 710. Page 700 depicted in Fig. 7A is merely illustrative of a multimedia paper document page according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[116] Page 700 depicted in Fig. 7A is imprinted with multimedia information corresponding to ten segments. According to an embodiment of the present invention, identifiers 712 identifying the segments are displayed in text sections 702 and 704, and in video section 708. The segment identifiers are printed proximally close to information corresponding to the respective segments. Page 700 also displays time span information 714 that indicates the start time and end time corresponding to information printed on page 700.

25 For example, the information printed on page 700 represents multimedia information recorded during the first 5:29 minutes of the recording. The page number 715 for each page is also displayed. Accordingly, page 700 depicted in Fig. 7A is the first page of the multimedia paper document.

[117] As shown in Fig. 7A, title section 702 displays title information for the 30 multimedia paper document. As depicted in Fig. 7A, the title information includes information identifying the source 716 of the multimedia information recording. According to an embodiment of the present invention, source information 716 corresponds to the name (e.g., filename) of the multimedia document for which the multimedia paper document has been generated. The title information displayed in section 702 also includes the time 718

when the multimedia document was recorded, the total time length 720 of the recording, and the date 722 of the recording. For example, page 700 is the first page from a multimedia paper document generated for “CNN News site (Channel 203)” television program which was recorded on May 30, 2001 starting at 12:59PM and has a total length of 56 minutes and

5 40 seconds.

[118] Text sections 704 and 706 display text information included in the multimedia document for which the multimedia paper document has been generated. In the embodiment depicted in Fig. 7A, text sections 704 and 706 display CC text included in the multimedia information. In alternative embodiments of the present invention, text sections
10 704 and 706 may display a transcript of the audio information included in the multimedia information.

[119] Identifiers 712 are printed in (or next to) to the text sections.

According to an embodiment of the present invention, each identifier 712 printed on a page of the multimedia paper document identifies a segment of the multimedia document that is printed on the page. The segment identifiers are printed proximally close to information corresponding to the respective segments.
15

[120] According to alternative embodiments of the present invention, identifiers 712 represent time points in the multimedia document. In this embodiment, identifiers 712 are printed proximal to information printed on a page that occurs close to the time corresponding to the identifier. For example, an identifier 712 printed on a particular page may represent a time in the time span for the particular page. For example, if the time span for a particular page is 0:00min-5:29min (e.g., time span of page 700 depicted in Fig. 7A), a particular identifier 712 may represent a time of 3:00min, i.e., 3 minutes into the multimedia recording. The particular identifier is printed proximal to information that occurs
20 at a time of 3 minutes into the multimedia recording.
25

[121] In the embodiment depicted in Fig. 7A, text sections 704 and 706 display the CC text in the English language. However, in alternative embodiments of the present invention that support multiple languages, the text may be printed in various languages or combinations thereof. The languages used to print the text may be different
30 from the language of the CC text included in the multimedia information or the language of the audio information included in the multimedia information. For example, the CC text associated with a video broadcast recording may be in English, but the text corresponding to the CC text printed on the multimedia paper document may be in a different language, for example, in Japanese (see Fig. 7D). Various different formats and styles may be used to print

text in the various languages. For example, according to an embodiment of the present invention, English text may be printed in text section 704 depicted in Fig. 7A and the corresponding translated Japanese text may be printed in text section 706. In alternative embodiments, each line of English text printed in a text section may be followed by a
5 Japanese translation of the text, and the like. Various other formats may also be used to print text in different languages. The translation of text from one language to another may be performed by MIPSS 104 or alternatively may be performed by some other service or application and then provided to MIPSS 104.

[122] The present invention also takes advantage of the automatic story
10 segmentation that is often provided in close-captioned (CC) text from broadcast news. Most news agencies who provide CC text as part of their broadcast use a special syntax in the CC text (e.g., a “>>>” delimiter to indicate changes in story line or subject, a “>>” delimiter to indicate changes in speakers, etc.) to indicate the beginning of a new story. Given the presence of this kind of information in the CC text transcript, the present invention can
15 further enhance the contents of the paper document with newspaper layout techniques such as bolding and line spacing that typically signify a new story. For example, as depicted in Fig. 7A, the first line of each new story is bolded. Further, additional spacing is provided between text portions related to different story lines to clearly demarcate the different stories. This further enhances the readability of the multimedia paper document.

[123] For each speaker identified in the CC text, information related to the speaker may be printed on page 700 (not shown). The information may include a name of the speaker, an address associated with the speaker, the title (e.g., CEO, etc.) of the speaker, and other like information related to or identifying the speaker. The information may also include information printed on a business card of the speaker. The information related to the
25 speakers may be determined from multimedia information stored in the multimedia document or may alternatively be determined from other information resources available to MIPSS 104.

[124] According to an embodiment of the present invention, video section 708 displays keyframes extracted from the video information corresponding to the CNN News Site (Channel 203) news recording. As depicted in Fig. 7A, four keyframes have been
30 extracted from the video information for each segment and displayed in video section 706. Identifiers 712 are printed in the upper right hand corner of each set of four keyframes. As described above, according to an embodiment of the present invention, identifiers 712 identify the segments from which the keyframes have been extracted. In alternative embodiments of the present invention, identifiers 712 may represent specific time points

within the multimedia recording. The number of keyframes that are extracted from each segment and the number of keyframes printed on each page of the multimedia paper document for each segment are user configurable. For example, according to one embodiment of the present invention, only one keyframe may be displayed for each segment, and the like. As previously stated, several different keyframe extraction techniques known to those of skill in the art may be used to extract keyframes from the video information included in the multimedia information. Additionally, several different techniques known to those of skill in the art may be used to display one or more of the extracted keyframes.

[125] As shown in Fig. 7A, identifiers 712 are printed in the text sections and also in the video section. A user may thus use identifiers 712 to correlate a portion of text printed in text sections 704 or 706 with a set of keyframes displayed in video section 708, and vice versa. For example, while a user is skimming the text section, the user may read a particular portion of text proximal to a particular identifier and locate keyframes related to or co-occurring with the particular portion of text using the particular identifier. Alternatively, the user may see an identifier for a particular keyframe (or set of keyframes) and use the identifier to locate text that describes what is being talked about at about the time that the keyframe(s) appeared in the video information. Identifiers 712 thus provide a sort of visual reference as well as a context for reading the text and the keyframes. This enhances the readability of the multimedia paper document.

[126] User-selectable identifiers 726 are printed on page 700. In the embodiment depicted in Fig. 7A, user-selectable identifiers 726 are printed as barcodes. A barcode 726 is printed for each segment printed on page 700. For example, barcode 726-1 corresponds to segment 1, barcode 726-2 corresponds to the segment 2, barcode 726-3 corresponds to the segment 3, and so on. In alternative embodiments of the present invention, various other techniques, besides barcodes, may be used to represent the user-selectable identifiers. As will be discussed below in further details, user-selectable identifiers 726 provide a mechanism for the reader of the multimedia paper document to access or retrieve multimedia information using the multimedia paper document.

[127] In alternative embodiments of the present invention where identifiers 712 represent specific time points in the multimedia information recording, barcodes 726 may be correlated to identifiers 712 and may also correspond to specific time points in the multimedia information recording. According to an embodiment of the present invention, barcodes 726 may correspond to the same time points as identifiers 712. Further details related to user-selectable identifiers 726 are provided below. User-selectable identifiers 726

are printed in a manner that does not reduce or affect the overall readability of the multimedia paper document.

[128] As depicted in Fig. 7A, controls section 710 displays a plurality of barcodes 724 corresponding to controls that may be used by a user to control playback of multimedia information corresponding to user-selected segments. Further details related to controls section 710 are provided below. Barcodes 724 are printed in a manner that does not reduce or affect the overall readability of the multimedia paper document.

[129] Fig. 7B depicts a second page 750 that follows page 700 depicted in Fig. 7A in a multimedia paper document according to an embodiment of the present invention. Title section 702 is not displayed on page 750. Page 750 displays text and keyframes corresponding to 11 segments (as compared to page 700 wherein information corresponding to 10 segments is displayed) of the multimedia document. The information displayed on page 750 corresponds to multimedia information corresponding to 5:29 minutes through 11:30 minutes of the recording (as indicated by time span information 714).

[130] Fig. 7C depicts a page 760 from a multimedia paper document generated according to an embodiment of the present invention for a multimedia document. Page 760 depicted in Fig. 7C corresponds to a page from a multimedia paper document generated for multimedia information recorded during a meeting. Information identifying the meeting is printed in title section 766. As depicted in Fig. 7C, page 760 comprises a first text section 762, a second text section 764, a video section 768, and a controls section 770.

[131] Closed-caption text (or a text transcript of the audio information) included in the multimedia document is printed in text section 762. A Japanese translation of the text printed in text section 762 is printed in text section 764. This is different from pages 700 and 750 depicted in Figs. 7A and 7B, respectively, wherein CC text was printed in both the text sections. For example, in Fig. 7A, the CC text printed text section 706 is a continuation of the text printed in text section 704. Various translation resources may be used to generate the Japanese translation printed in section 764 of Fig. 7C. It should be apparent that in alternative embodiments, the CC text may be translated to other languages and printed in a multimedia paper document.

[132] Page 760 depicted in Fig. 7C is merely illustrative of a multimedia paper document page according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[133] Given a multi-paged multimedia paper document comprising pages of the type depicted in Figs. 7A, 7B, or 7C, a reader can quickly skim the contents of the multimedia paper document to see if anything relevant might be present in the multimedia information for which the multimedia paper document was generated. The time required to skim and comprehend information printed in the multimedia paper document will be much smaller than the time the user would otherwise have to spend viewing the multimedia information (e.g., new broadcast recording). The present invention thus allows the user to save valuable time when “reading” multimedia information.

[134] Fig. 8A depicts a page 800 from a multimedia paper document generated for a recorded meeting according to an embodiment of the present invention. Page 800 depicted in Fig. 8A is merely illustrative of a page from a multimedia paper document and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[135] The recorded meeting for which page 800 is generated may store multimedia information that includes video information, audio information, slides information, and whiteboard information. Techniques for recording meetings has been described in U.S. Non-Provisional Patent Application No. 09/728,560, filed November 30, 2000, and U.S. Non-Provisional Patent Application No. 09/728,453, filed November 30, 2000.

[136] The slides information included in a recorded meeting may comprise information related to slides (e.g., a PowerPoint presentation slides) presented by the presenter during the meeting. The whiteboard information may comprise information related to text and drawings drawn on a whiteboard during the meeting. Accordingly, in addition to text (which may correspond to a transcript of the audio information) and video information, slides information and whiteboard information are also included in the printable representation of the recorded meeting. The text, video, slides, and whiteboard information may then be printed on a paper medium as depicted in Fig. 8. Accordingly, the text information is printed in sections 802 and 804, video information is printed in section 806, and slides 806 and whiteboard images 808 are printed inline with the text sections.

[137] According to an embodiment of the present invention, during generation of the printable representation for the recorded meeting, MIPSS 104 synchronizes the slides information and whiteboard information with the audio and video information using timestamps associated with the various types of information. When the multimedia information corresponding to the recorded meeting is divided into segments, each segment

may comprise text information, video information, slides information, and whiteboard information. When the multimedia paper document is generated, one or more slides are then printed in close proximity to the identifier of a segment that contains slides information related to the printed slides. The slides are thus printed close to when they were presented.

5 Likewise, images of the whiteboard are printed in close proximity to the identifier of a segment that contains the whiteboard information. The whiteboard images are thus printed close to when they were presented. In the embodiment depicted in Fig. 8A, the slides and whiteboard images are printed inline with the text sections.

[138] Various different layout and format guidelines may be used for
10 printing the slides and whiteboard information. For example, in Fig. 8B, slides 808 and whiteboard images 810 are printed in the margins of the multimedia paper document next to text sections 802. Fig. 8C shows yet another layout pattern for printing the slides and whiteboard information. In Fig. 8C, the slides and whiteboard images are superimposed on video keyframes belonging to the segments to which the slides and whiteboard images belong.

[139] As described above, audio information included in multimedia information stored by a multimedia document is displayed in the form of a text transcript of the audio information. According to an embodiment of the present invention, various other features of the audio signal included in the multimedia information may also be represented in the printable representation of the multimedia document. According to an embodiment of the present invention, visual markers are used to represent the various features and when printed on a paper medium improve the overall readability and understandability of the multimedia paper document.

[140] For example, Fig. 9A depicts a page 900 from a multimedia paper
25 document displaying visual markers to denote various attributes of the audio information or of the CC text information included in the multimedia information for the multimedia document for which the multimedia paper document is generated according to an embodiment of the present invention. Page 900 depicted in Fig. 9A is merely illustrative of a multimedia paper document page according to an embodiment of the present invention and
30 does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[141] As depicted in Fig. 9A, a gap or white space 908 is shown in text section 904 corresponding to one more segments that do not contain any CC text information but may comprise other types of information (e.g., video information) which is printed on the

page (e.g., keyframes in video section 906). The gap may represent a section of the recording wherein there is no audio or CC text. Alternatively, the gap may represent a section of the recording where there is no CC text and the audio information during the section cannot be translated to text. For example, someone is speaking in a foreign language for which an

5 English translation is not available. The length of gap 908 may be proportional to the length of the empty CC text or absence of audio information. If the multimedia information does not include both the audio information and the CC text for a specified time period, a visual marker such as "SILENCE" may be printed in gap 908.

[142] The video information during a gap in the CC text or audio information
10 may still contain important information and is thus displayed as keyframes in video section
906. For example, if someone is speaking in a foreign language for which an English
translation is not available for display in the text section, the video during this period may
display text (e.g., subtitles) corresponding to what is being spoken. Accordingly, keyframes
displaying the text may be printed in video section 906 while a gap is printed in the text
section. According to one embodiment of the present invention, the text images may be
extracted from the video keyframes during the gap period and printed in gap 908. For
example, as depicted in Fig. 9B, text images 920 have been extracted from video keyframes
corresponding to the gap period, and the extracted text images 920 are printed in gap space
908. According to yet another embodiment of the present invention, optical character
recognition (OCR) techniques may be applied to the video keyframes for the gap period and
the results of the OCR may be printed in gap space 908. For example, as depicted in Fig. 9C,
OCR techniques have been applied to the video keyframes during the gap period, and the
resultant OCRed text 930 (which may contain spelling errors) is printed in gap 908.

[143] Other features of the audio information may also be represented via
25 visual markers printed on the multimedia paper document. For example, features of audio
information such as people singing, multiple people talking at the same time, people arguing,
speaking in soothing tones, significant increases in audio volumes, periods of silence
(described above) etc. can be identified and represented in the multimedia paper document
using visual markers. For example, as depicted in Fig. 9A, visual markers "(Singing)" 910
30 are printed where the audio information contains people singing. The visual markers thus
make it easy for the reader of the multimedia paper document to quickly locate such parts of
the audio in the multimedia document.

[144] Several different techniques known to those skilled in the art may be used to identify special features of the audio information in the multimedia information. The

following references discuss a few techniques that may be used to identify features of audio signals. The entire contents of the following references are herein incorporated by reference for all purposes:

(1) L.S. Chen, H. Tao, T.S. Huang, T. Miyasato, R. Nakatsu, "Emotion
5 Recognition from Audiovisual Information," Proc. IEEE Workshop on Multimedia Signal Processing, Los-Angeles, CA, USA, pp. 83-88, 1998;

(2) K. Sonmez, L. Heck, M. Weintraub, "Multiple Speaker Tracking and Detection: Handset Normalization and Duration Scoring," Digital Signal Processing, 10(1/2/3), 133-143, 2000; and

10 (3) F. Dellaert, T. Polzin, A. Waibel, "Recognizing emotion in speech." Proceedings ICSLP 96. Fourth International Conference on Spoken Language Processing (Cat. No.96TH8206). IEEE. Vol.3, pp.1970-1973, 1996. New York, NY, USA.

[145] As described above, video information included in multimedia information stored by a multimedia document is displayed in the form of one or more keyframes extracted from the video information and printed on the multimedia paper document. According to an embodiment of the present invention, various other features of the video information included in the multimedia information may also be represented in the printable representation of the multimedia document. According to an embodiment of the present invention, visual markers may be used to represent the various features of the video information and when printed on a paper medium improve the overall readability and understandability of the multimedia paper document.

[146] For example, features that can be recognized from video information may include faces, facial expressions of speakers depicted in the video (e.g., a facial expression indicating anger), recognition of speakers, hand gestures, logos or signs displayed in the video, presence of certain buildings or geographical locations, meetings, animals, crowds, and the like. According to an embodiment of the present invention, these features are recognized from the video information and represented in the multimedia paper documents using visual markers. For example, expressions (e.g. "Anger," "Laughter", etc.), geographical locations, special building names, etc. can be shown with a text-based annotation next to the corresponding video keyframe. Speaker face recognition results may be shown with the name of the speaker printed next to a keyframe depicting the speaker. Logos and signs recognized from the video can be displayed or the names of companies corresponding to the logos may be displayed.

[147] Several different styles and formats may be used to display the visual markers for the video features. According to an embodiment of the present invention, the layout and format information may specify the styles or formats are to be used. For example, the visual markers for the video information may be displayed close to keyframes

5 corresponding to the visual markers, or in the margins of the multimedia paper document. In alternative embodiments, the visual markers may be displayed interspersed with the text section information. Different types of fonts may be used to display the visual markers. The visual markers thus make it easy for the reader of the multimedia paper document to quickly locate such parts of the video in the multimedia document.

10 [148] The text displayed in the text sections of a multimedia paper document (e.g., text sections 704 and 706 depicted in Fig. 7A) may also be modified based upon recognition of certain video features. For example, text printed in the text section and spoken by a particular speaker may be highlighted, and the like.

15 [149] Several different techniques known to those skilled in the art may be used to identify special features of the video information in the multimedia information. The following references discuss a few techniques that may be used to identify features of the video data. The entire contents of the following references are herein incorporated by reference for all purposes:

20 (1) A. Essa, A. P. Pentland, Coding Analysis Interpretation and Recognition of Facial Expressions, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757-763, 1997;

(2) G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, Classifying Facial Actions, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, Oct. 1999;

25 (3) A.F. Bobick, A.D. Wilson, A State based approach to the representation and recognition of gesture, IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 1325-1337, 1997;

30 (4) H. A. Rowley, S. Baluja, T. Kanade, "Neural network-based face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 20, no. 1, 23-38, 1998;

(5) D. S. Doermann, E Rivlin, and I. Weiss. Applying algebraic and differential invariants for logo recognition. Machine Vision and Applications, 9(2):73-86, 1996;

(6) H. Li, D. Doermann, and O. Kia. Automatic Text Detection and Tracking in Digital Video. *IEEE Transactions on Image Processing - Special Issue on Image and Video Processing for Digital Libraries*, 9(1), pages 147-156, 2000;

5 (7) P. Suda, C. Bridoux, B. Kammerer, G. Manderlechner, "Logo and word matching using a general approach to signal registration," *Fourth International Conference on Document Analysis and Recognition*, Ulm, Germany, August 18-20, 1997, 61-65;

(8) H. Li, D. Doermann, and O. Kia. Text Extraction and Recognition in Digital Video. *Proceedings of Third IAPR Workshop on Document Analysis Systems*, pages 119-128, 1998;

10 (9) Face recognition techniques described at web site "www.visionics.com";
and

(10) Ioffe, S.I. and Forsyth, D.A., Finding people by sampling, Proc. International Conference on Computer Vision, p.1092-7, 1999.

15 [150] Various other features of the multimedia information may also be detected and represented in the printable representation of the multimedia document (or on the multimedia paper document when printed), using special visual markers. For example, the presence of a commercial in the multimedia information may be detected and information corresponding to the commercial printed on the paper medium (e.g., keyframes corresponding to the commercial, portions of text sections corresponding to the commercial, etc.) may be visually demarcated (e.g., by using a special font, drawing boxes around the printed information, etc.). As another example, sections of the multimedia information including multiple individuals talking within a user-configurable length of time may be identified and specially marked in the multimedia paper document. For example, a user may like to see parts of the multimedia information where more than 3 different people speak
20 within a 1-minute period. This information may be highlighted in the multimedia paper document.

25 [151] Several different techniques known to those skilled in the art may be used to identify special features of the video information in the multimedia information. The following reference discusses a few techniques that may be used to identify features of the video data. The entire contents of the following reference are herein incorporated by reference for all purposes.

30 (a) Rainer Lienhart, Christoph Kuhmünch and Wolfgang Effelsberg. On the Detection and Recognition of Television Commercials, Proc. IEEE Conf. on Multimedia Computing and Systems, Ottawa, Canada, pp. 509 - 516, June 1997.

[152] ANNOTATING MULTIMEDIA INFORMATION

[153] According to the teachings of the present invention, the printable representation for a multimedia document may be annotated to identify locations of information in the multimedia document that may be of interest to the user. The multimedia paper document generated by printing the annotated printable representation on a paper medium displays the annotations. The annotations provide visual indications of information relevant or of interest to the user. For example, information printed in the multimedia paper document that is relevant to topics of interest specified by a user may be annotated or highlighted. In this manner, the multimedia paper document provides a convenient tool that allows a user to readily locate portions of the multimedia paper document that are relevant to the user. Since the multimedia paper document comprises a printable representation of multimedia information, the multimedia paper document generated according to the teachings of the present invention allows the user to identify portions of multimedia information that are of interest to the user.

15
16
17
18
19
20
21
22
23
24
25

25

30

[154] According to an embodiment of the present invention, information specifying topics of interest to the user may be stored in a user profile. One or more words or phrases may be associated with each topic of interest. Presence of words and phrases associated with a particular user-specified topic of interest indicates presence of information related to the particular topic. For example, a user may specify two topics of interest--“George Bush” and “Energy Crisis”. Words or phrases associated with the topic “George Bush” may include “President Bush,” “the President,” “Mr. Bush,” and other like words and phrases. Words or phrases associated with the topic “Energy Crisis” may include “industrial pollution,” “natural pollution,” “clean up the sources,” “amount of pollution,” “air pollution”, “electricity,” “power-generating plant,” and the like. Probability values may be associated with each of the words or phrases indicating the likelihood of the topic of interest given the presence of the word or phrase. Various tools may be provided to allow the user to configure topics of interest, to specify keywords and phrases associated with the topics, and to specify probability values associated with the keywords or phrases.

[155] According to an embodiment of the present invention, after generating a printable representation of multimedia information stored in a multimedia document (in step 408 of Fig. 4), MIPSS 104 accesses the user profile information and determines topics of interest specified in the user profile and keywords and phrases associated with the topics of interest. MIPSS 104 then searches the printable representation of the multimedia information

to identify locations within the printable representation of words or phrases associated with the topics of interest. As described above, presence of words and phrases associated with a particular user-specified topic of interest indicates presence of the particular topic relevant to the user. According to one embodiment of the present invention, MIPSS 104 searches the
5 text sections included in the printable representation of a multimedia document to locate words or phrases associated with the user topics. If MIPSS 104 finds a word or phrase in the printable representation that is associated with a topic of interest, the word or phrase is annotated in the printable representation. Several different techniques may be used to annotate the word or phrase. For example, the word or phrase may highlighted, bolded,
10 underlined, demarcated using sidebars or balloons, font may be changed, etc. The annotations are then printed on the multimedia paper document generated by printing the annotated printable representation of the multimedia information.

[156] According to an embodiment of the present invention, MIPSS 104 may also highlight keyframes (representing video information of the multimedia document) related to user specified topics of interest. According to an embodiment of the present invention, MIPSS 104 may use OCR techniques to extract text from the keyframes included in the printable representation of the multimedia information. The text output by the OCR techniques may then be compared with words or phrases specified in a user's profile. If there is a match, the keyframe corresponding to the matched word or phrase (i.e., the keyframe from which the matching word or phrase was extracted) may be annotated in the printable representation. Several different techniques may be used to annotate the keyframe. For example, a special box may surround the keyframe, the matching text in the keyframe may be highlighted or underlined or displayed in reverse video, and the like. The keyframe annotations are then printed on the multimedia paper document generated by printing the
25 annotated printable representation of the multimedia information.

[157] According to another embodiment of the present invention, MIPSS 104 may identify information stored by the multimedia document that is relevant to user-specified topics of interest even before the printable representation for the multimedia document has been generated. In this embodiment, MIPSS 104 analyzes the multimedia information stored
30 in the multimedia document to identify information relevant to user-specified topics of interest. For example, MIPSS 104 may analyze the video information contained in the multimedia document to identify video frames that contain information relevant to user-specified topics of interest. Various different techniques, e.g., OCR techniques, known to those skilled in the art may be used to analyze the video information. MIPSS 104 may

analyze the audio or closed-caption text information included in the multimedia document to identify sections of the information that include information relevant to user-specified topics of interest. For example, MIPSS 104 may generate a text transcript of the audio information and then analyze the text transcript to identify presence of words or phrases related to the
5 user-specified topics of interest. Likewise, the CC text may also be analyzed. Other types of information (e.g., slides information, whiteboard information, etc.) included in the multimedia information stored by the multimedia document may also be analyzed. As previously stated, various analysis techniques known to those skilled in the art may be used to analyze the multimedia information stored by the multimedia document. MIPSS 104 may
10 then generate a printable representation for the multimedia document and annotate information in the printable representation that was deemed relevant to one or more user-specified topics of interest. The multimedia paper document generated by printing the annotated printable representation displays the annotations.

15 [158] Fig. 10 depicts a page 1000 whose contents have been annotated according to an embodiment of the present invention. Page 1000 depicted in Fig. 10 is merely illustrative of a multimedia paper document page and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. As depicted in Fig. 10, words and phrases related to topics of interest are highlighted in text sections 1002 and 1004. For the embodiment
20 depicted in Fig. 10 it is assumed that two topics of interest, namely “George Bush” and “Energy Crisis”, have been specified. Keywords and phrases related to these topics of interest are highlighted. Different colors and styles (e.g., bolding, underlining, different font size, etc.) may be used to highlight words and phrases related to different topics. For example, as depicted in Fig. 10, a first color is used to highlight words and phrases related to the “George Bush” topic of interest and a second color is used to highlight words and phrases related to the “Energy Crisis” topic of interest.
25

30 [159] According to an embodiment of the present invention, in addition to highlighting information relevant to topics of interest, the present invention may also determine and display a relevancy score for each topic of interest. The relevancy score calculated for a particular topic of interest indicates the relevancy of the information printed in the multimedia paper document to that particular user topic. The relevancy score for a topic of interest thus indicates the degree of relevancy of the multimedia information represented by the multimedia paper document to the topic of interest. According to an embodiment of the present invention, the relevancy score for a particular topic may be

calculated based upon the frequency of occurrences of the words and phrases associated with the particular topic in the multimedia paper document.

[160] The relevancy scores for the topics may be included in the printable representation of the multimedia document and printed on the multimedia paper document.

- 5 A reader or user could then use the relevancy score printed on the multimedia paper document as a guide to determine relevancy of the multimedia information to the user topics. For example, if multiple multimedia paper documents have been generated for a plurality of news broadcasts, based upon the relevancy scores printed on the multimedia paper documents for the various broadcasts, a user can easily determine the news broadcast that is most
- 10 relevant to the user for any given user topic.

[161] According to an embodiment of the present invention, information stored in a user's profile and words or phrases related to user-specified topics of interest detected in the text section (CC text or transcript of audio information) may also be used to select keyframes from the video information that are relevant to the user-specified topics of interest. Since only a limited number of keyframes can be printed on the multimedia paper document due to limited amount of space available on a multimedia paper document page, selection of keyframes relevant to the user improves the readability of the document for the user.

[162] As described above, a user profile may be configured by a user and may store information identifying one or more topics of interest to the user. One or more words or phrases may be associated with each topic of interest such that presence of the words and phrases associated with a particular topic of interest indicates presence of information related to the particular topic. According to an embodiment of the present invention, probability values may be associated with each of the words or phrases indicating the likelihood of the topic of interest given the presence of the word or phrase. In order to facilitate selection of relevant keyframes, the user profile also stores information about features of the video information (or of the keyframes) that the user would like the present invention to search for in the video information when a word or phrase related to a topic is found in the text section.

30 [163] As previously described, several different features can be recognized from the video information. These features may include recognition of a human face, buildings, geographical locations, presence of a crowd, hand gestures, logos or signs, meetings, animals, text, and the like. Various algorithms known to those skilled in the art may be used to detect the features from the video information. For each of the features stated

above, techniques that recognize video features may also be used to identify specific instances of a feature. For example, if a face is identified in a video frame, face recognition techniques may be applied to determine the identity of the face (i.e., a specific instance of a face). Likewise, if a logo was identified in a video frame, techniques may be applied to

5 determine the name of the company corresponding to the logo. Similarly, if a building was identified in a video frame, techniques may be applied to determine if the building was a specific building such as the Empire State Building. Likewise, if an animal was identified in a video frame, techniques may be applied to determine the type (e.g., horse, cat, dog, etc.) of animal.

10 [164] As part of a user's profile, the user may specify one or more video features (and specific instances of the features where appropriate) to be associated with one or more topics of interest. According to an embodiment of the present invention, the video features may be associated with keywords or phrases associated with user topics of interest. For each video feature, the user may also specify weight values for each topic indicating the relative importance of the video feature for that topic of interest.

15 [165] Fig. 11 depicts a user profile 1100 that may be configured by a user according to an embodiment of the present invention to facilitate selection of keyframes relevant to user-specified topics of interest. As depicted in Fig. 11, three topics of interest have been specified, namely, "Terrorism", "Company XYZ", and "Football". Keywords and phrases have been associated with each of the topics. In order to facilitate selection of keyframes relevant to the user topics of interest from the video information, the user has also specified video features to be searched for when keywords and phrases associated with the topics are located in the text (e.g., in the CC text or in the text transcript of audio information) of the multimedia information. Weights have been associated with the video features
20 indicating the relative importance of the video features for each topic of interest. For example, for the topic "Terrorism", the face of Osama Bin Laden (weighted 0.7) is slightly more important than presence of text "Afghanistan" (weighted 0.6).

25 [166] Profile 1100 specifies the criteria for selecting keyframes relevant to the topics of interest given the presence of a keyword or phrase related to the topics of interest. For example, profile information 1100 specifies that the words "Osama" and "Afghanistan" are associated with the topic "Terrorism". If the word "Osama" is located in the text information of the multimedia information, then the video information (or video frames which have been extracted from the video information) temporally proximal to the occurrence of the word "Osama" in the text information are to be checked to determine if

they include a face of Osama Bin Laden. Keyframes that contain Osama Bin Laden's face are deemed to be relevant (degree of relevance indicated by the weight value) to topic "Terrorism."

[167] Likewise, if the word "Afghanistan" is located in the text information of the multimedia information, then the video frames temporally proximal to the occurrence of the word "Afghanistan" in the text information are to be checked to determine if they contain text "Afghanistan". As previously described, OCR techniques may be used to extract text from video keyframes. Keyframes that contain text "Afghanistan" are deemed to be relevant (degree of relevance indicated by the weight value) to topic "Terrorism."

[168] Further, for all (indicated by "*") keywords and phrases (including "Osama" and "Afghanistan") associated with the topic "Terrorism," video frames temporally proximal to the occurrence of the words or phrases in the text information are to be checked to determine if they contain a building or (indicated by the Boolean connector OR) a crowd. Such keyframes are deemed to be relevant (degree of relevance indicated by the weight value) to topic "Terrorism." Accordingly, if the word "Osama" is located in the text information of the multimedia information, then the video frames temporally proximal to the occurrence of the word "Osama" in the text information would be first checked to determine if they include a face of Osama Bin Laden, and then checked to determine if they contain a building or a crowd.

[169] Likewise, profile information 1100 specifies that the word "Suzuki" is associated with the topic "Company XYZ" ("Suzuki" may be the name of the CEO of Company XYZ). If the word "Suzuki" is located in the text information of the multimedia information, then the video frames temporally proximal to the occurrence of the word "Suzuki" in the text information are to be checked to determine if they include a face of John Suzuki. Keyframes that contain John Suzuki's face are deemed to be relevant (degree of relevance indicated by the weight value) to topic "Company XYZ."

[170] Further, for all (indicated by "*") keywords and phrases (including "Suzuki") associated with the topic "Company XYZ", video frames temporally proximal to the occurrence of the words or phrases are to be checked to determine if they contain a building and (indicated by the Boolean connector AND) further if they contain either a XYZ logo or (indicated by the Boolean connector OR) text "XYZ". Such keyframes are deemed to be relevant (degree of relevance indicated by the weight value) to topic "Company XYZ."

[171] Likewise, profile information 1100 specifies that the phrase "Buffalo Bills" is associated with the topic "Football". If the phrase "Buffalo Bills" is located in the

text information of the multimedia information, then the video frames temporally proximal to the occurrence of the phrase are to be checked to determine if they include a face of Jim Kelly or the face of Marv Levy. Keyframes that contain either Jim Kelly's or Marv Levy's face are deemed to be relevant (degree of relevance indicated by the weight value) to topic

5 "Football".

[172] Fig. 12 depicts modules that facilitate selection of keyframes relevant to topics of interest according to an embodiment of the present invention. The modules depicted in Fig. 12 may be software modules, hardware modules, or combinations thereof. The modules depicted in Fig. 12 are merely illustrative of an embodiment of the present

10 invention and are not meant to limit the scope of the present invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[173] As depicted in Fig. 12, a video feature recognition module 1202 receives as input video frames corresponding to (or extracted from) the video information contained by a multimedia document. For each video frame, video feature recognition module 1202 determines if the video frame contains any video features such as a face, a building, a logo, text, etc. If a particular video feature is located, video feature recognition module 1202 assigns a probability value to the feature indicating the probability that the video frame contains the identified video feature. Video feature recognition module 1202 also determines a specific instance of the video feature and assigns a probability value to it. For example, for a particular video frame, video feature recognition module 1202 may determine that there is an 85% probability that the video frame contains a face and that there is a 90% probability that the face belongs to John Suzuki. For the same video frame, video feature recognition module 1202 may determine that there is only a 3% probability that the

20 video frame contains a building, and only a 1% probability that the video frame contains a logo. The output of video feature recognition module 1202 is a ranked list of features and specific instances for each video frame. If no video feature is detected, a generic keyframe selection procedure may be applied. The procedure may calculate the probability that a frame is a potential keyframe. The video frames and their associated ranked list information

25 30 is then forwarded to frame selector module 1204 for further processing.

[174] Profile matcher module 1206 receives as input user profile information and text information (e.g., CC text information or transcript of audio information) extracted from the multimedia document. Based upon the user profile information, profile matcher module 1206 searches the text information to locate words and phrases in the text information

that are related to user-specified topics of interest. The words and phrases located in the text information are annotated by profile matcher module 1206. As described previously, various techniques may be used to annotate the text information. The text information along with the annotations is then forwarded to frames selector module 1204 and to printable representation generator module 1208 for further processing.

[175] As described above, frame selector module 1204 receives as input video frames and associated ranked list information from video feature recognition module 1202, the annotated text information from profile matcher module 1206, and user profile information. Based upon the user profile information, for each annotated word or phrase found in the text information, frame selector module 1204 determines video frames relevant to the topic with which the annotated word or phrase is associated.

[176] According to an embodiment of the present invention, for a particular annotated word, frame selector module 1204 identifies video frames temporally proximal to the occurrence of the particular annotated word in the text information. This may be performed by determining the time stamp associated with the annotated word, and identifying a set of video frames within a time window of $\pm N$ seconds from the time stamp of the annotated word. The value of N is user configurable. For example, if $N=5$ seconds, frame selector module 1204 identifies a set of video frames within a time window of ± 5 seconds from the time stamp of the annotated word.

[177] Frame selector module 1204 then determines the topic of interest with which the annotated word is associated and the video features that are relevant to the particular annotated word or topic of interest (as specified in the user profile). Each video frame in the set of video frames within the $\pm N$ seconds time window is then searched to determine if it contains one or more video features specified in the user profile for the topic or word. A relevancy score is calculated for each video frame in the set of video frames.

[178] According to an embodiment of the present invention, in order to calculate a relevancy score for a video frame, frame selector module 1204 multiplies the weight assigned to the video feature in the user profile by the probability value assigned to the video frame by video feature recognition module 1202 indicating that the frame contains the particular video feature. The other probability values assigned to the video frame by video feature recognition module 1202 may be multiplied by a constant (K) that is less than the weights in the profile information. This ensures that the simultaneous detection of a keyword and a relevant video frame will provide a higher rank for that video frame than if the keyword was not detected. After each video frame in the set of video frames has been

assigned a relevancy value, the video frames are ranked based upon their relevancy values. Accordingly, for each annotated word or phrase in the text information, frame selector module generates a ranked list of video frames. The ranked list of keyframes for each annotated word or phrase is then forwarded to printable representation generator module

5 1208 for further processing.

[179] Printable representation generator module 1208 receives annotated text information from profile matcher module 1206 and ranked lists of keyframes for the annotations from frame selector module 1204. Printable representation generator module also receives as input other types of information included in the multimedia information

10 stored by the multimedia document and layout and format information. Based upon the various inputs, printable representation generator module 1208 generates a printable representation for the multimedia document.

[180] According to an embodiment of the present invention, as part of processing performed to generate the printable representation for the multimedia document, printable representation generator module 1208 determines which keyframes to be included in the printable representation for the multimedia document for each segment of the multimedia document based upon the layout and format information and the ranked listing received from frame selector module 1204. For example, let's assume that the layout information specifies that four keyframes are to be printed for each segment. In this scenario, if the text corresponding to a particular segment contains only one annotated word, then the top four most relevant keyframes from the ranked list of keyframes associated with the annotated word and received from frame selector module 1204 are selected to be included in the printable representation for the segment. If a particular segment contains four different annotated words, then printable representation generator module 1208 may select only the most relevant keyframe from the ranked lists of keyframes associated with each of the four annotated words for inclusion in the printable representation (making a total of 4 keyframes). Accordingly, printable representation generator module 1208 determines the keyframes to be included in the printable representation for each segment of the multimedia document using the ranked list of keyframes received from frame selector module 1208.

30

[181] USING A MULTIMEDIA PAPER DOCUMENT TO RETRIEVE MULTIMEDIA INFORMATION

[182] The present invention provides techniques that allow a user to access or retrieve multimedia information in digital form using the multimedia paper document

generated for a multimedia document. The multimedia paper document may thus be used as an indexing and retrieval tool for retrieving multimedia information that may be stored in the multimedia document. For example, a user may use a multimedia paper document generated for a video recording to access or retrieve portions of the video recording.

5 [183] Fig. 13A is a simplified high-level flowchart 1300 depicting a method of retrieving multimedia information using a multimedia paper document according to an embodiment of the present invention. Flowchart 1300 depicted in Fig. 13 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

10 [184] As depicted in Fig. 13, a user may initiate the method by selecting one or more segments from a multimedia paper document corresponding to multimedia information that the user wishes to access or retrieve (step 1302). The segments may be selected by selecting user-selectable identifiers (e.g., user-selectable identifiers 726 depicted in Fig. 7A) associated with the segments using a selection device. The user-selectable identifiers corresponding to the segments may be selected from one or more pages of a multimedia paper document. Further, the user-selectable identifiers may be selected from one or more multimedia paper documents. Several different techniques may be provided by a multimedia paper document to enable the user to select one or more segments.

15 [185] According to an embodiment of the present invention, barcode technology is used to facilitate selection of segments. In this embodiment, the user-selectable identifiers are printed in the multimedia paper document in the form of barcodes. Each barcode corresponds to a particular segment of the multimedia document. For example, as depicted in Fig. 7A, according to an embodiment of the present invention, a barcode 726 is printed for each segment printed on a page of the multimedia paper document. For example, barcode 726-1 printed on page 700 corresponds to segment 1, barcode 726-2 corresponds to the segment 2, barcode 726-3 corresponds to the segment 3, and so on. A user can select a particular segment by scanning the barcode corresponding to that segment. A selection device such as a barcode scanner or any other device that is capable of scanning barcodes 20 may be used to scan the barcodes. The user may scan one or more barcodes from one or more pages of one or more multimedia paper documents.

25 [186] It should be apparent that various other techniques, besides barcodes, may be used to facilitate selection of segments corresponding to multimedia information that the user wishes to access or retrieve. According to an embodiment of the present invention,

the user-selectable identifiers may be implemented as watermarks printed on pages of the multimedia paper document may also be used as techniques for selecting segments. In this embodiment, a user may select one or more watermarks corresponding to segments of interest to the user using a selection device that is capable of reading or detecting the watermarks.

5 [187] According to another embodiment of the present invention, the user-selectable identifiers may be implemented as text string identifiers printed on pages of the multimedia paper document. In this embodiment, a user may select a particular segment by keying in or entering the text string identifier corresponding to the particular segment into a selection device such as a telephone, a DVR, etc.

10 [188] Various other techniques (e.g., Xerox glyphs embedded in keyframes, etc.) known to those skilled in the art may also be used to facilitate selection of segments. Generally, in order to maintain the readability of the multimedia paper document, techniques that are less obtrusive, and those that do not take up too much space on the page, and which are somewhat aesthetically pleasing may be used.

15 [189] After the user has selected one or more segments, the user may select preferences for playing back the multimedia information corresponding to the segments selected in step 1302 (step 1304). According to an embodiment of the present invention, the user may specify preferences by selecting one or more controls from controls section 710. As with selection of segments, various different techniques maybe used to facilitate selection of controls. For example, according to an embodiment of the present invention, a particular control may be selected by scanning a barcode corresponding to the control. For example, the user may specify that the multimedia information is to be played back in “Enhanced Mode” by selecting barcode 724-4 depicted in Fig. 7A. The user may specify that the playback is to show CC text by selecting barcode 724-5 corresponding to control “Show 20 Closed-caption”. The user may specify that time is to be displayed during the playback by selecting barcode 724-6 corresponding to control “Show Time”. The user in step 1304 may also select various other preferences.

25 [190] According to an embodiment of the present invention, as part of step 1304, the user may also specify an output device to be used for playing back the multimedia 30 information corresponding to the segments selected in step 1302. According to an embodiment of the present invention, one or more devices that may be located at different geographical locations may be selected for playback. For example, the selected output device may be the user’s PDA, a computer in the user’s office, a television at the user’s home, a specific kiosk, and the like.

[191] In alternative embodiments of the present invention, the user preferences and the output device may be pre-configured. For example, this information may be stored in a user profile. Alternatively, the preferences and the output device may default to some preset values. In such a scenario, step 1304 may not be performed.

5 [192] The user may then request playback of the multimedia information corresponding to the segments selected in step 1302 (step 1306). According to an embodiment of the present invention, the user may request playback by selecting a barcode such as barcode 724-1 corresponding to the “Play” control. According to an embodiment of the present invention, upon selecting the “Play” control, a signal is transmitted from the
10 selection device (e.g., a barcode scanner) used by the user to select the segments and the preferences to a server that is capable of retrieving multimedia information corresponding to the user-selected segments. The server may be MIPSS 104 or any other server. The signal communicated to the server from the selection device may identify the segments selected by the user in step 1302, the multimedia paper documents from which the segments are to be selected, information related to preferences and/or one or more output devices selected by the user in step 1304, and other like information to facilitate retrieval of the requested multimedia information.
T5

20 [193] Various techniques and communication links may be used to communicate the signal from the selection device used by the user to the server. For example, if the selection device is a barcode scanner, a communication link may be established between the barcode scanner and the server and the signal information may be communicated to the server via the communication link. Different types of communication links may be used including hardwire links, optical links, satellite or other wireless communications links, wave propagation links, or any other mechanisms for communication
25 of information. Various communication protocols may be used to facilitate communication of the signal via the communication links. These communication protocols may include TCP/IP, HTTP protocols, extensible markup language (XML), wireless application protocol (WAP), protocols under development by industry standard organizations, vendor-specific protocols, customized protocols, and others.

30 [194] In other embodiments, a telephone may be used as a selection device. For example, a user may use a telephone to establish a communication link with the server. The user may then communicate the signal information to server using the telephone. For example, the user may key in user-selectable identifiers (e.g., text string identifiers)

corresponding to the selected segments and controls using the telephone. Various other techniques may also be used to communicate the information to the server.

[195] The server receiving the signal from the selection device may then retrieve multimedia information corresponding to the user-selected segments (step 1308).

5 According to an embodiment of the present invention, the server determines the user-selectable identifiers selected by the user. The server then determines segments of the multimedia document corresponding to the user-selectable identifiers selected by the user. The server then retrieves multimedia information corresponding to the selected segments.

[196] The multimedia information may be retrieved from a single 10 multimedia document or from multiple multimedia documents. For example, if the user selected user-selectable identifiers from multiple multimedia documents, then the server retrieves multimedia information corresponding to selected segments from multiple multimedia documents.

[197] The multimedia information retrieved by the server is then 15 communicated to the one or more output devices selected for playback (step 1310). The multimedia information is then played on the one or more output devices selected for playback (step 1312). The user may control playback of the multimedia information by selecting one more controls from control area 710 depicted in Fig. 7A. For example, the user may stop playback of the multimedia information by selecting barcode 724-1 corresponding to the “Stop” control. A user may fast-forward 10 seconds of the multimedia information by selecting barcode 724-2. A user may rewind 10 seconds of the multimedia information by selecting barcode 724-3. Various other controls not shown in Fig. 7A may also be provided 20 in alternative embodiments of the present invention to control playback of the multimedia information.

25 [198] According to an alternative embodiment of the present invention, a user may use the multimedia paper document to start playing multimedia information from a user-selected time point in the multimedia document. In this embodiment, the user-selectable identifiers (e.g., barcodes 726 depicted in Fig. 7A) printed in a multimedia paper document represent particular time points in the multimedia document. According to an embodiment of 30 the present invention, the barcodes may correspond to the same time points as the identifiers (e.g., identifiers 712 depicted in Fig. 7A) printed on a page of the multimedia paper document.

[199] Fig. 13B is a simplified high-level flowchart 1350 depicting a method of retrieving multimedia information from a particular time point using a multimedia paper

document according to an embodiment of the present invention. Flowchart 1350 depicted in Fig. 13B is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

5 [200] As depicted in Fig. 13, a user may initiate the method by selecting a user-selectable identifier printed on a multimedia paper document corresponding to a time point in the multimedia document from where the user wishes to retrieve multimedia information (step 1352). As described above, several different techniques (e.g., barcodes, watermarks, glyphs, text strings, etc.) may be provided by the multimedia paper document to enable the user to the user-selectable identifier.

10 [201] After selecting a user-selectable identifier, the user may select preferences for playing back the multimedia information (step 1304). As described above with respect to Fig. 13A, the user may select a mode for playing back the multimedia information, select one or more output devices for playing back the multimedia information, and the like. Step 1354 may not be performed if the user preferences are pre-configured.

15 [202] The user may then request playback of the multimedia information (step 1356). According to an embodiment of the present invention, upon selecting the “Play” control, a signal is transmitted from the selection device (e.g., a barcode scanner) used by the user to a server that is capable of retrieving multimedia information from the multimedia document. The server may be MIPSS 104 or any other server. The signal communicated to the server from the selection device may identify the user-selectable identifier selected by the user in step 1352, the multimedia paper document from which the user-selectable identifier was selected, information related to preferences and/or one or more output devices selected by the user in step 1354, and other like information to facilitate retrieval of the requested

20 multimedia information.

25 [203] The server receiving the signal from the selection device then retrieves multimedia information from the time point corresponding to the user-selectable identifier selected by the user in step 1352 (step 1358). According to an embodiment of the present invention, the server determines a time point in the multimedia document corresponding to the user-selectable identifier selected by the user and retrieves multimedia information from the time point onwards. view

30 [204] The multimedia information retrieved by the server in step 1358 is then communicated to the one or more output devices selected for playback (step 1360). The multimedia information is then played back on the one or more output devices selected for

playback (step 1362). The user may control playback of the multimedia information by selecting one or more controls from control area 710 depicted in Fig. 7A. For example, the user may stop playback of the multimedia information by selecting barcode 724-1 corresponding to the “Stop” control. A user may fast-forward 10 seconds of the multimedia information by 5 selecting barcode 724-2. A user may rewind 10 seconds of the multimedia information by selecting barcode 724-3. Various other controls not shown in Fig. 7A may also be provided in alternative embodiments of the present invention to control playback of the multimedia information.

[205] Accordingly, as described above, the multimedia paper document 10 provides a simple and easy-to-use mechanism for retrieving multimedia information. The convenience afforded by the multimedia paper document in retrieving multimedia information might be illustrated by the following example. Let's assume that a user has requested that the television program “Bloomberg” be recorded between the hours of 9-11AM during which important financial news is broadcast. Various different devices may be used to record the news broadcast including a video recorder, a digital video recorder (DVR) (e.g., a TIVO box), and the like. The user may then generate a multimedia paper document 15 for the recorded news broadcast.

[206] Let's further assume that the user has 15 minutes before a power lunch with a client to digest the two-hour Bloomberg TV program to find out if any relevant 20 information was mentioned regarding the client's company or their main competitor. With the paper-based version of the broadcast (i.e., the multimedia paper document), the user can quickly skim the multimedia paper document for relevant information. When the user finds one or more segments in the multimedia paper document of interest, the user can use a barcode scanner to scan the barcodes corresponding to segments in the multimedia paper 25 document. The user may also scan a control barcode instructing the recorder to launch the video corresponding to the selected segments on a television in the user's office. This sends a message to the recorder to launch the video corresponding to the selected segments on the television in the user's office. If the user has selected multiple segments, multimedia 30 information corresponding to the selected segments will be played on the user's television, skipping the segments or sections that are not selected by the user. In this manner, the user can quickly navigate two-hours of a television broadcast in 15 minutes watching only those portions of the broadcast that are of interest to the user, and be ready for the client meeting in time.

[207] In the above scenario the user could have selected segments from multiple multimedia paper documents generated for a plurality of news broadcasts from news agencies such as CNBC, CNN/fn, MSNBC, and the like. The user could then skim the multimedia paper documents to locate news related to the client or the client's competitors from the various broadcasts. This is equivalent to watching several hours of video in a short time—something that is very difficult to achieve if the user only has access to a video player. The user may then select segments of interest from the multiple multimedia papers and watch video corresponding to the selected segments.

[208] In the above scenario, the present invention automatically records a desired broadcast program based a user's profile and produces a multimedia paper document that acts both as a familiar skimming tool and a retrieval device for viewing desired portions of the video. In the above-described scenario, the interface is not on the user's personal computer--instead, the interface is in the user's hands in the form of paper. In some cases, this is a more desired environment since most individuals are familiar with and more comfortable with reading and using paper. The paper-based interface thus provides a unique mechanism for indexing or referring back to the digitized multimedia information stored by the multimedia document. The indexing technique provided by the present invention may then be used by a user to retrieve the multimedia information in digital format. The multimedia paper document provides a portable means for random access to the multimedia information, a task that traditionally required tedious searching of the multimedia information.

[209] GENERATING A SINGLE PRINTABLE REPRESENTATION FOR
A PLURALITY OF MULTIMEDIA DOCUMENTS

[210] The present invention provides techniques for generating a single printable representation that includes multimedia information extracted from a plurality of different multimedia documents or multimedia sources. According to an embodiment of the present invention, the single printable representation includes multimedia information selected from the plurality of multimedia documents based upon selection criteria. A user may specify the selection criteria. The selection criteria may be based upon any attributes of the multimedia documents or their contents, or upon user-specified topics of interest, and the like. For example, the selection criteria may specify a particular subject (e.g., information related to the Taliban in Afghanistan, or abortion related information, etc.), a specified story line, and the like.

[211] For example, a user may specify that a single printable representation (or a single multimedia paper document) be generated consolidating stories and articles related to “Middle East Terrorism” from a plurality of news broadcast recordings. In response, the present invention generates a single printable representation that includes 5 multimedia information from the plurality of news broadcast recordings related to “Middle East Terrorism.” The single consolidated printable representation may then be printed to generate a single consolidated multimedia paper document that contains information related to “Middle East Terrorism” from multiple multimedia documents.

[212] According to another example, topics of interest to the user (which 10 may be stored in a user profile) may be specified as the selection criteria. Based upon such selection criteria, MIPSS 104 may generate a single printable representation that includes multimedia information from the plurality of news broadcast recordings related to the user-specified topics of interest. The single consolidated printable representation may then be 15 printed to generate a single consolidated multimedia paper document that contains information related to “Middle East Terrorism” extracted from multiple multimedia documents. In this manner, multimedia information from various multimedia sources or 20 documents related to user-specified topics of interest may be consolidated into a single printable representation that may then be printed to generate a multimedia paper document. The multimedia paper document generated in this manner is a valuable tool that enables the user to read and comprehend related information from multiple sources in a timely and efficient manner.

[213] Fig. 14 is a simplified high-level flowchart 1400 depicting a method of 25 generating a single printable representation according to an embodiment of the present invention that includes multimedia information selected from a plurality of multimedia documents by analyzing the printable representations of the plurality of multimedia documents. The method depicted in Fig. 14 may be used to generate a single multimedia paper document including multimedia information selected from a plurality of multimedia documents. The processing depicted in Fig. 14 may be performed by software modules executing on MIPSS 104, by hardware modules coupled to MIPSS 104, or a combination 30 thereof. In alternative embodiments of the present invention, the processing may be distributed among the various systems depicted in Fig. 1. The processing depicted in Fig. 14 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[214] The method is initiated by determining the selection criteria (or criterion) to be used for selecting the multimedia information to be included in the single printable representation and by determining the plurality of multimedia documents (or multimedia sources) from which the multimedia information is to be selected or extracted
5 (step 1402). MIPSS 104 then generates a printable representation for each multimedia document determined in step 1402 if a printable representation does not already exist for the multimedia document (step 1404). The printable representations for the multimedia documents may be generated according to the methods depicted in Figs. 4 and 6.

[215] For each multimedia document identified in step 1402, MIPSS 104
10 searches the pages from the printable representation of the multimedia document to identify a set of pages that comprise information that satisfies the selection criteria determined in step 1402 (step 1406). MIPSS 104 then generates a single consolidated printable representation that includes the pages determined in step 1406 (step 1408). The single printable representation generated in step 1408 may then be printed on a paper medium to generate a consolidated multimedia paper document (step 1410). The multimedia paper document generated in step 1410 comprises information selected from the plurality of multimedia documents based upon the selection criteria. For each page of the multimedia paper document generated in step 1410, information printed information that satisfies the selection criteria may be annotated.
15

[216] As described above, the printable representations of the multimedia documents are analyzed to identify portions of multimedia information from the various multimedia documents to be included in the consolidated printable representation. According to alternative embodiments of the present invention, the multimedia information stored by the multimedia documents may be analyzed to identify portions of the multimedia information
20 that satisfy the selection criteria. A consolidated printable representation may then be generated to include portions of multimedia information from the various multimedia documents that satisfy the selection criteria. The consolidated printable representation may then be printed on a paper medium to generate a consolidated or “customized” multimedia paper document.

[217] Fig. 15 is a simplified high-level flowchart 1500 depicting another method of generating a single printable representation that includes information extracted from a plurality of multimedia documents by analyzing the multimedia information stored by the plurality of multimedia documents according to an embodiment of the present invention. The method depicted in Fig. 15 may be used to generate a single multimedia paper document
30

including multimedia information extracted from a plurality of multimedia documents. The processing depicted in Fig. 15 may be performed by software modules executing on MIPSS 104, by hardware modules coupled to MIPSS 104, or a combination thereof. In alternative embodiments of the present invention, the processing may be distributed among the various systems depicted in Fig. 1. The processing depicted in Fig. 15 is merely illustrative of an embodiment incorporating the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[218] The method is initiated by determining the selection criteria (or criterion) to be used for selecting the multimedia information to be included in the single printable representation and by determining the plurality of multimedia documents (or multimedia sources) from which the multimedia information is to be selected (step 1502). For each multimedia document determined in step 1502, MIPSS 104 divides the multimedia information contained by the multimedia document into segments of a particular time length (step 1504). The process of dividing a multimedia document into segments has been described earlier with respect to Figs. 6.

[219] For each multimedia document identified in step 1502, MIPSS 104 then determines those segments or portions of the multimedia document that comprise information that satisfies the selection criteria identified in step 1502 (step 1506). MIPSS 104 then generates a single consolidated printable representation based upon the segments determined in step 1506 (step 1508). The single consolidated printable representation includes segments determined in step 1506. The single printable representation generated in step 1508 may then be printed on a paper medium to generate a consolidated multimedia paper document (step 1510). The multimedia paper document generated in step 1510 comprises information selected from the plurality of multimedia documents based upon the selection criteria. The multimedia paper document generated in step 1510 may comprise annotations identifying printed information that satisfies the selection criteria.

[220] A multimedia paper document generated according to the flowcharts depicted in Figs. 14 and 15 may then be used as any other multimedia paper document. For example, a user may select one or more user-selectable identifiers from the consolidated multimedia paper document (as described above) and retrieve multimedia information corresponding to segments associated with the user-selectable identifiers selected by the user.

[221] Figs. 16A, 16B, 16C, and 16D depict pages of a multimedia paper document generated according to an embodiment of the present invention using the method

depicted in Fig. 14. The pages have been selected from a plurality of multimedia documents because they contain information related to the topic of interest “Middle East Terrorism” that was specified as the selection criteria. The pages have been selected from printable representations generated for a plurality of multimedia documents. For example, pages 1600
5 and 1602 depicted in Figs. 16A and 16B have been selected from a printable representation generated for a “CNN News Site (Channel 203)” recording that was recorded on May 30, 2001 starting at 12:59PM and is of length 56:40 minutes, page 1606 depicted in Fig. 16C has been selected from a printable representation generated for a “Newshour (PBS, Channel 233)” recording that was recorded on June 5, 2001 starting at 6:01PM and is of length 54:49
10 minutes, page 1604 depicted in Fig. 16D has been selected from a printable representation generated for a “Hardball (CNBC, Channel 356)” recording that was recorded on September 14, 2001 starting at 5:00PM and is of length 59:59 minutes. For each page, information related to “Middle East Terrorism” has been annotated. This enhances the readability of the multimedia paper document. Accordingly, information related to “Middle East Terrorism” from a plurality of multimedia documents is consolidated into one document.
15

[222] As described above, a user may generate a “customized” multimedia paper document by specifying appropriate selection criteria. In this manner, the user can quickly extract relevant information from multiple hours of multimedia broadcasts by simply reading the customized multimedia paper document. The present invention thus reduces the time spent by the user in locating and retrieving relevant information from multiple multimedia information sources or recordings.
20

[223] COVERSHEETS

[224] According to an embodiment of the present invention, the present
25 invention also provides techniques for generating a coversheet for a multimedia paper document. The coversheet may provide a summary of the contents printed in the multimedia paper document.

[225] Fig. 17 depicts a coversheet 1700 generated for a multimedia paper document according to an embodiment of the present invention. Coversheet 1700 depicted in
30 Fig. 17 is merely illustrative of a coversheet according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[226] As shown in Fig. 1700, coversheet 1700 comprises thumbnail images of individual pages included in the multimedia paper document. As shown, eight thumbnail

images 1704 are printed on coversheet 1700 thereby indicating that the multimedia paper document comprises eight pages. A title section 1702 is also printed on coversheet 1700. Title section 1702 displays the source of the multimedia information (which may correspond to the filename of the multimedia document), the time and date when the multimedia information was recorded, and the total time of the recording.

5 [227] A barcode 1706 is associated with each thumbnail image. A user may use barcodes 1706 to access or retrieve multimedia information printed on the pages in digital form. If the user wishes to access the multimedia information corresponding to information printed on a particular multimedia paper document page, the user may scan the barcode
10 corresponding to that particular page and then access or retrieve the information in electronic form using an output device specified by the user. In this embodiment, selecting a barcode associated with a particular page is equivalent to selecting all the segments printed on that particular page. For example, if the user wishes to access multimedia information corresponding to the information printed on page 6 of the multimedia paper document, the user may scan barcode 1706-6 and then access the information (as previously described) using an output device. The user may select one or more barcodes from coversheet 1700.

15
20
25

[228] According to another embodiment of the present invention, a barcode 1706 associated with a particular page is the same as the barcode corresponding to the first segment printed on the particular page. In this embodiment, the user may scan a barcode for a particular page and retrieve multimedia information starting from the top of the particular page.

25 [229] Fig. 18 depicts a coversheet 1800 generated for a multimedia paper document according to another embodiment of the present invention. In addition to the features included in coversheet 1700 depicted in Fig. 17, coversheet 1800 displays a list of sentences 1804 for each thumbnail image 1802. According to an embodiment of the present invention, the sentences displayed for a particular thumbnail image summarize the contents of the page corresponding to the particular thumbnail image. Several different techniques may be used to select the sentences for a particular thumbnail image. According to an embodiment of the present invention, the first text sentence of each segment printed on the 30 page corresponding to the thumbnail image may be printed in 1804. According to another embodiment of the present invention, segments that contain CC text with story-line separators (e.g., ">>>"), the first sentence of each story printed on the page corresponding to the thumbnail image may be printed in 1804. Other techniques known to those skilled in the art may also be used to determine the text to be printed in area 1804 of coversheet 1800.

[230] It should be apparent that coversheet 1800 depicted in Fig. 18 is merely illustrative of a coversheet according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

5 [231] Fig. 19 depicts a coversheet 1900 generated according to another embodiment of the present invention for a multimedia paper document that has been annotated based upon user-specified topics of interest. A title section 1902 is printed on coversheet 1900 displaying the source of the multimedia information (which may correspond to the filename of the multimedia document), the time and date when the multimedia 10 information was recorded, and the total time of the recording. Topics of interest 1904 to which the multimedia paper document is relevant are also displayed. For each topic of interest, the degree of relevancy of the multimedia paper document to the topic of interest is also displayed. In the embodiment depicted in Fig. 1900, the degree or relevancy is denoted by a percentage value 1906.

15 [232] Coversheet 1900 displays a thumbnail image 1908 of each page included in the multimedia paper document. For pages that comprise information related to user-specified topics, the thumbnail images corresponding to those pages display the annotated words or phrases related to user-specified topics of interest. For a particular page comprising information related to one or more user-specified topics of interest, the number of 20 hits 1910 related to the topics of interest found on the particular page are also displayed next to the thumbnail image of the page. Different colors and styles may be used to highlight words and phrases in the thumbnails related to different topics. The hits for a particular topic of interest may also be displayed using a color that is associated with the topic of interest and used to highlight words and phrases related to the topic of interest. This allows the user of 25 the multimedia paper document to easily identify pages of the multimedia paper document that include information related to user-specified topics of interest.

30 [233] It should be apparent that coversheet 1900 depicted in Fig. 19 is merely illustrative of a coversheet according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[234] Fig. 20 depicts a coversheet 2000 generated according to an embodiment of the present invention for a multimedia paper document that includes pages selected from multiple multimedia paper documents based upon selection criteria. For example, the multimedia paper document may be generated according to flowchart 1400

depicted in Fig. 14. Coversheet 2000 depicted in Fig. 20 has been generated for a multimedia paper document that includes pages 1600, 1602, 1604, and 1606 depicted in Figs. 16A, 16B, 16C, and 16D, respectively. It should be apparent that coversheet 2000 depicted in Fig. 20 is merely illustrative of a coversheet according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[235] As depicted in Fig. 20, the selection criteria 2002 used for generating the multimedia paper document is printed on page 2000. Coversheet 2000 displays a thumbnail image 2004 of each page included in the multimedia paper document. For pages 10 that comprise information related to the search criteria, the thumbnail images corresponding to those pages displaying the information with annotations. The number of hits 2006 for pages are also displayed. A barcode 2008 associated with each page is also displayed. Coversheet 2000 also displays a date range 2010 that may be selected by the user as part of the selection criteria. For example, multimedia paper document comprises information in the date range May 1, 2001 to September 20, 2001.

[236] Fig. 21 depicts another coversheet 2100 generated according to an embodiment of the present invention for a multimedia paper document that includes pages selected from multiple multimedia paper documents based upon selection criteria. Coversheet 2100 depicted in Fig. 21 has been generated for a multimedia paper document 15 that includes pages 1600, 1602, 1604, and 1606 depicted in Figs. 16A, 16B, 16C, and 16D, respectively. It should be apparent that coversheet 2000 depicted in Fig. 20 is merely 20 illustrative of a coversheet according to an embodiment of the present invention and does not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives.

[237] Coversheet 2100 shows more information than coversheet 2000 depicted in Fig. 2000. For each occurrence of words or phrases related to the selection criteria (e.g., text related to “Middle East Terrorism”), the line 2102 (or a user-configurable 25 number of words surrounding the relevant word/phrase) comprising the relevant text or phrase (which is annotated) is displayed along with the time 2104 when the word/phrase occurred in the recording and the page 2106 of the multimedia paper document on which the line is printed.

[238] A barcode 2108 is also displayed for each line. According to an embodiment of the present invention, barcode 2108 corresponds to the barcode for the page on which the line occurs. According to alternative embodiments of the present invention, the

barcode 2108 associated with a line may correspond to the barcode of the segment that contains the displayed line. Alternatively, barcode 2108 may correspond to a location within the multimedia information when the relevant text/phrase occurs. Accordingly, barcode 2108 enables the user to access or retrieve multimedia information from a specific point in time.

5 [239] A set of keyframes 2110 is also displayed for each line. According to an embodiment of the present invention, the keyframes that are most representative of the word/phrase or are relevant to the selection criteria may be displayed. Techniques for selecting keyframes relevant to selection criteria such as a user-specified topic of interest have been described above.

10 [240] Fig. 22 depicts a coversheet 2200 generated according to an embodiment of the present invention for a multimedia paper document that has been generated for a recorded meeting. As shown in Fig. 22, coversheet 2200 comprises thumbnail images of individual pages included in the multimedia paper document. As shown, six thumbnail images 2202 are printed on coversheet 2200 thereby indicating that the multimedia paper document comprises eight pages. A title section 2204 is also printed on coversheet 2200 and displays information identifying the meeting for which the multimedia paper document was generated, the time and date when the meeting was recorded, and the total time of the recording. Slides 2206 and whiteboard images 2208 are also printed next to thumbnail images corresponding to pages that comprise the slides or whiteboard images.

15 [241] It should be apparent that coversheets 1700, 1800, 1900, 2000, 2100, and 2200 depicted in Figs. 17, 18, 19, 20, 21, and 22, respectively, are merely illustrative of specific embodiments of the present invention and do not limit the scope of the invention as recited in the claims. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The coversheet generated according to the teachings of the
20 present invention thus provide a simple and convenient way for the reader of the multimedia paper document to get an overview of the contents of the multimedia paper document.

25 [242] Although specific embodiments of the invention have been described, various modifications, alterations, alternative constructions, and equivalents are also encompassed within the scope of the invention. The described invention is not restricted to operation within certain specific data processing environments, but is free to operate within a plurality of data processing environments. Additionally, although the present invention has been described using a particular series of transactions and steps, it should be apparent to

those skilled in the art that the scope of the present invention is not limited to the described series of transactions and steps.

[243] Further, while the present invention has been described using a particular combination of hardware and software, it should be recognized that other 5 combinations of hardware and software are also within the scope of the present invention. The present invention may be implemented only in hardware, or only in software, or using combinations thereof.

[244] The specification and drawings are, accordingly, to be regarded in an 10 illustrative rather than a restrictive sense. It will, however, be evident that additions, subtractions, deletions, and other modifications and changes may be made thereunto without departing from the broader spirit and scope of the invention as set forth in the claims.